Reinhard Schrank

A Theoretical Analysis of Strategic Auditor-Client Interaction

Dissertation

to be awarded the degree of

Doctor of Social and Economic Sciences (Dr. rer.soc.oec.)

at the University of Graz

First Supervisor:  Univ.-Prof. Dr. Ralf Ewert,
Department of Accounting and Auditing

Second Supervisor:  O.Univ.-Prof. Dr. Dr.hc. Alfred Wagenhofer,
Department of Accounting and Control

Graz, February 2019
# Contents

**Acknowledgments**  
IV  

**Author’s Declaration**  
V  

**List of Abbreviations**  
VII  

**List of Figures**  
VIII  

## 1 Introduction  
1  

1.1 Outline  
1  

1.2 State of Research  
10  

1.3 Research Questions and Approach  
18  

1.3.1 Main Research Question  
18  

1.3.2 Dissertation Papers  
21  

1.4 Methodology  
24  

## 2 The Impact of Damage Apportionment on ICS Quality and Financial Reporting Accuracy  
26  

2.1 Introduction  
27  

2.2 Related Literature  
30  

2.3 The Model  
33  

2.4 The Equilibrium  
41  

2.5 Comparative Statics  
44  

2.5.1 ICS Quality  
44  

2.5.2 Audit Quality  
46  

2.5.3 Financial Reporting Accuracy  
49  

2.5.4 Social Welfare  
52  

2.6 Robustness of the Results  
56
# TABLE OF CONTENTS

2.7 Conclusion ................................................................. 60
2.8 Appendix A ................................................................. 62
2.9 Appendix B ................................................................. 77
    2.9.1 Outcome of the ICS is observable before the auditor is hired . . . 77
    2.9.2 Investors receive less than full loss compensation ................. 83

3 Audit Quality, Legal Liability, and the Audit Market under Risk Aversion 86
    3.1 Introduction ........................................................... 87
    3.2 Related Literature .................................................... 89
    3.3 Analysis ................................................................. 92
        3.3.1 Audit Quality under Risk Aversion ............................... 92
        3.3.2 The Audit Market under Risk Aversion ......................... 95
        3.3.3 Effects of Liability Regulation .................................... 105
    3.4 Regulatory Implications and Conclusion ............................. 119
    3.5 Appendix ............................................................... 122

4 Sequential Auditor-Client Interaction under Strategic Effort Complementarity 136
    4.1 Introduction ........................................................... 137
    4.2 Empirical Evidence .................................................... 141
    4.3 The Model .............................................................. 144
    4.4 The Equilibrium ....................................................... 151
    4.5 Comparative Statics ................................................... 154
        4.5.1 Internal Effort ...................................................... 154
        4.5.2 Audit Effort ...................................................... 160
        4.5.3 Financial Reporting Accuracy ................................... 162
        4.5.4 Heterogeneous Stakeholder Interests ......................... 164
Acknowledgments

Writing this dissertation would not have been possible without the support of many people. To some of these people I would like to express my particular gratitude.

First, I thank my main thesis supervisor and department head, Univ.-Prof. Dr. Ralf Ewert for sparking my interest in academic research, giving me the possibility to pursue a dissertation project as a research assistant at the Department of Accounting and Auditing at the University of Graz, and for his continuous encouragement and support in my research endeavors. I also thank the faculty members O. Univ.-Prof. Dr. Dr. h.c. Alfred Wagenhofer, Univ.-Prof. Dr. Rainer Niemann, Univ.-Prof. Dr. Dr. Georg Schneider, Univ.-Prof. Christoph Kuzmics, PhD and Ao. Univ.-Prof. Dr. Heinz Königsmaier for their frequent and valuable advice, and for providing a fruitful research environment. Special thanks go to Professor Wayne Landsman for having me as a visiting scholar at the Kenan-Flagler Business School of the University of North Carolina at Chapel Hill in the fall semester 2017. These two months have been an invaluable experience, both from an academic and personal perspective.

I am also grateful to my fellow PhD students and colleagues, who made these four years in the PhD program a wonderful and rewarding time. In particular, I thank Birgit Beinsen, Robert Eder, Sebastian Kronenberger, Claudia Langbauer, Felix Niggemann, Elisabeth Plietzsch, and David Windisch for their valuable input. A special thanks goes to Harald Hinterecker for the many thoughtful discussions we had, but also for the many joyful academic and non-academic events we shared. I also thank Elke Efferl for her help and support during the nearly seven years I worked at the department.

Above all, I thank my parents Elisabeth and Josef for encouraging my interest in education from my earliest childhood on, and for their strong support in every aspect of my life. I also thank my girlfriend Marlen Rieger for her support and patience, as well as my close friend Peter Pfleger for his ongoing friendship.

I gratefully acknowledge the financial support from the Austrian Science Fund (FWF), project W-1229.
Author’s Declaration

Unless otherwise indicated in the text or references, or acknowledged above, this thesis is entirely the product of my own scholarly work. Any inaccuracies of fact or faults in reasoning are my own and accordingly I take full responsibility. This thesis has not been submitted either in whole or part, for a degree at this or any other university or institution. This is to certify that the printed version is equivalent to the submitted electronic one.

Date:

(Reinhard Schrank)
List of Abbreviations

2SLS Two-stage least squares
ARA Absolute risk aversion
AS5 Auditing Standard No. 5
CDFC Convexity of the distribution function condition
CEO Chief executive officer
CPA Certified public accountant
Eds. Editors
e.g. example given
et seq. et sequentes
FOA First-order approach
FOC First-order condition
GAAP Generally Accepted Accounting Principles
GAAS Generally Accepted Auditing Standards
IAS International Accounting Standard
ICS Internal control system
i.e. id est
LHS Left hand side
MAR Missing at random
MICE Multiple imputation by chained equations
MLRC Monotone likelihood-ratio condition
NAS Non-audit services
NPV Net present value
OLS Ordinary least squares
PCAOB Public Company Accounting Oversight Board
PSLRA Private Securities Litigation Reform Act of 1995
p page
pp pages
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;D</td>
<td>Research and development</td>
</tr>
<tr>
<td>RHS</td>
<td>Right hand side</td>
</tr>
<tr>
<td>SEC</td>
<td>Securities and Exchange Commission</td>
</tr>
<tr>
<td>SOX</td>
<td>Sarbanes-Oxley Act of 2002</td>
</tr>
<tr>
<td>S&amp;P</td>
<td>Standard and Poor's</td>
</tr>
<tr>
<td>St. Dev.</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>s.t.</td>
<td>subject to</td>
</tr>
<tr>
<td>U.S.</td>
<td>United States</td>
</tr>
<tr>
<td>VNL</td>
<td>Vague negligence regime</td>
</tr>
<tr>
<td>VNM</td>
<td>Von Neumann-Morgenstern</td>
</tr>
<tr>
<td>Vol.</td>
<td>Volume</td>
</tr>
<tr>
<td>w.l.o.g.</td>
<td>without loss of generality</td>
</tr>
<tr>
<td>w.r.t.</td>
<td>with respect to</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Financial accounting and auditing in a capital market context</td>
<td>10</td>
</tr>
<tr>
<td>1.2</td>
<td>State of research on the auditor’s contribution to accurate financial reporting</td>
<td>17</td>
</tr>
<tr>
<td>1.3</td>
<td>Contributions of the dissertation</td>
<td>25</td>
</tr>
<tr>
<td>2.1</td>
<td>The model economy</td>
<td>33</td>
</tr>
<tr>
<td>2.2</td>
<td>Game tree of the model</td>
<td>35</td>
</tr>
<tr>
<td>2.3</td>
<td>Equilibrium levels of ICS quality $s^*$</td>
<td>47</td>
</tr>
<tr>
<td>2.4</td>
<td>Equilibrium levels of audit effort $a^*$</td>
<td>48</td>
</tr>
<tr>
<td>2.5</td>
<td>Equilibrium social welfare levels $U^{SW}(a^b, s^b)$ and $U^{SW}(a^<em>, s^</em>)$</td>
<td>55</td>
</tr>
<tr>
<td>2.6</td>
<td>Game tree of the model with observable ICS outcome</td>
<td>78</td>
</tr>
<tr>
<td>3.1</td>
<td>Auditors’ effort choice $a^*$ for different degrees of risk aversion $\gamma$</td>
<td>104</td>
</tr>
<tr>
<td>3.2</td>
<td>Distribution of surplus $S$</td>
<td>104</td>
</tr>
<tr>
<td>3.3</td>
<td>Market equilibria for $H$ and $H'$</td>
<td>105</td>
</tr>
<tr>
<td>3.4</td>
<td>Expected utility curves for the logarithmic specification with parameters $y = 0.05, p = 0.9, k = 1.75, V \in {15, 17, 19}$</td>
<td>107</td>
</tr>
<tr>
<td>4.1</td>
<td>The financial reporting process</td>
<td>149</td>
</tr>
<tr>
<td>4.2</td>
<td>The exponential audit technology $d(a, s) = 1 - e^{sa}$ for different levels of internal effort $s$</td>
<td>154</td>
</tr>
<tr>
<td>4.3</td>
<td>The square root audit technology $d_m(a, s) = \sqrt{sa}$ for different levels of internal effort $s$</td>
<td>159</td>
</tr>
<tr>
<td>4.4</td>
<td>Equilibria for the exponential model specification</td>
<td>165</td>
</tr>
<tr>
<td>4.5</td>
<td>Equilibria for the square root model specification</td>
<td>165</td>
</tr>
<tr>
<td>4.6</td>
<td>Equilibria for the multiplicative model specification</td>
<td>166</td>
</tr>
<tr>
<td>4.7</td>
<td>Sequence of events in the model with heterogeneous stakeholder interests</td>
<td>168</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Outline

Financial accounting is an information system that is designed to produce, distribute and process information among market participants in order to protect individuals and markets from the detrimental consequences of asymmetric information. Asymmetric information means that one market participant is better informed about certain aspects of a situation than one or more other market participants. "More" or "better" information is typically interpreted as an information system $S'$ being sufficient for another information system $S$ in the sense that $S'$ can be obtained from $S$ by adding noise. For discrete information systems, this is means that each signal is "finer", in that it is systematically linked with less states of the world. The concepts of sufficiency and fineness of information systems have been developed by Blackwell (1951) and Blackwell / Girshik (1954).

Asymmetric or incomplete information is part of a vicious circle that starts and ends with imperfect and incomplete markets. A market is perfect if there are no trade impediments in the form of entry-/exit-barriers, taxes, transaction or information costs, and each market participant maximizes his or her expected utility based on the same information and beliefs. On a complete market the number of tradable claims equals the number of states of nature. On a perfect and complete capital market each and every market participant would trade any claim for future cash flows as long as his or her marginal utility derived from the cash flow stream is maximized. The market price that emerges in equilibrium will be a compound of all market participants’ relative marginal utilities, incorporating each individual’s preferences and information endowment. Under these conditions, the value of
a claim is well-defined and equals the observable market price. However, in reality markets are neither perfect nor complete, which means that not all utility-maximizing trades that market participants would prefer to undertake, actually take place. Therefore not all information and preferences of each and every market participant enter the market price, and value is no longer well-defined. In other words, the asymmetric information about certain trade-relevant aspects translates into asymmetric information about the value of a claim. Then, market participants who possess more information can use this advantage to make bargains at the cost of the less informed. In general this causes inefficient resource allocation, because either the worse informed individual enters trade and gets exploited (given individual rationality, this would happen if the worse informed party is unaware of the information asymmetry), or the worse informed party refrains from trade at all in order to protect him- or herself from being exploited. The latter not only implies allocation inefficiency among individuals, but can lead to the failure of whole markets, as famously demonstrated by [Akerlof (1970)].

In the context of financial accounting the better informed "insider" is typically the firm’s management, whereas the less informed "outsiders" span the wide range of the firm’s stakeholders. The main markets a firm participates in are the capital market, the labor market and the product market. The primary goal of financial accounting is to mediate inefficiencies that arise from information asymmetries at the capital market. Hence, the main firm stakeholders addressed by financial accounting are the firm’s present and potential future equity- and debtholders. Of course, also labor market participants (current and prospective employees), and product market participants (suppliers, customers, competitors) use financial accounting information. However, they are not the first addressees of financial accounting.

Situations with asymmetric information are most commonly classified according to (i) what kind of information is asymmetrically distributed, and (ii) when the information asymmetry arises. The first type of information asymmetry is hidden characteristics.
INTRODUCTION

Hidden characteristics means that the outsider is to some degree uninformed about certain characteristics of the firm (e.g. the quality of the firm’s investment portfolio, or the ability of its managers) and thus refers to information asymmetry that is present prior to trade. If the contract between in- and outsider is made, outsiders may select a type they would not have selected if they were informed about the characteristics. This type of problem is called adverse selection. Asymmetric information problems that arise after trade happens are hidden action, hidden information and hidden intention. Hidden action means that the outsider is ignorant about the insider’s action. Under hidden information the outsider knows the insider’s action but does not understand its implications. In both cases the outsider can only observe the final outcome, but never knows what action the insider did in fact take. This informational advantage gives rise to opportunistic incentives of the insider, so called moral hazard. For example, a firm might obtain external funding for an investment project. However, once having received the funds, managers might immediately use them for private consumption, leaving the new investors with an empty shell of a firm. It is clear that ex ante an investor who is aware of those incentives would not be willing to provide capital without being able to protect him- or herself from being expropriated. Finally, hidden intention means that the outsider is uninformed about the insider’s goals even though he or she might be able to observe the insider’s action. For example, a manufacturer and a supplier enter a relationship which requires a specific investment by the supplier. After the contract is made, the manufacturer reduces the wholesale price for the input good which hurts the supplier. The supplier can perfectly observe this behavior but is forced to accept the lower price because the specific investment cannot (or only at high costs) be used otherwise. Ex ante the supplier will therefore not undertake the investment.

Financial accounting tackles the goal of overcoming or at least mitigating the problems of information asymmetry at the lowest costs to the economy in two distinct but related ways. First and most obvious, accounting should reduce the information asymmetry itself
by transmitting information from the insider to the outsider. This allows the outsider to make better economic decisions, since he or she can better assess the suitability of economic transactions to maximize his or her utility. For example, if reported earnings are informative about future dividend payouts, then investors can use them to decide whether they are better off in buying, holding, or selling claims against the firm. In a wider sense that includes nature as a market participant, also the gathering of new information can be understood as reduction of information asymmetry. Improving the decision-making of market participants by reducing asymmetric information is the first main role of financial accounting (decision-making role). Second, whenever information asymmetry cannot be fully reduced, financial accounting should limit the possibilities for better informed insiders to take advantage on the less informed outsiders, and thereby enable, respectively facilitate efficiency-enhancing transactions. This is achieved by contractually making financial claims for cash flows contingent on accounting information in a way such that the insiders’ and the outsider’s incentives are aligned. For example, manager’s incentives to maximize their private consumption can be aligned with shareholders’ interest in maximizing the stream of dividends by making both claims contingent on the firm’s reported earnings. Facilitating transactions in the presence of asymmetric information is the second main role of financial accounting (contracting role).

This dissertation focuses on reduction of information asymmetry through financial accounting, that is its decision-useful role. In order to be decision-useful, information must be relevant and reliable. Relevance requires that the information is suitable to change the recipient’s expectations about past, current or future states of the world. For example, reported earnings are relevant to investors if they are in some way linked to fu-

---

1 Many of the attributes that make financial accounting decision-useful also facilitate contracts that use accounting information (e.g. veracity, verifiability). Hence, the results of the dissertation can also be interpreted from the contracting perspective. However, depending on the contract some decision-useful attributes may be suboptimal from a contracting perspective. This is compactly demonstrated in Wagenhofer / Ewert (2014) with the example of fair value- versus historical cost accounting; see Wagenhofer / Ewert (2014), pp 189 et seq.
ture dividends, because then earnings information changes the expectations about future consumption potential. Reliability requires that the information accurately describes, respectively predicts past, current or future states of the world, that is recipients of the information are right in believing that the described or predicted states of the world in fact occur. To stick with the earnings example, an earnings figure that is based on the possible but extremely unlikely default of a key customer with historically excellent creditworthiness is not reliable. Rational, aware investors should not adapt their expectations about the firm’s future ability to pay dividends, even though the default of the customer would entail severe consequences.

The natural way to reduce information asymmetry at capital markets and to improve the individual decision-making of market participants is full disclosure of relevant and reliable information by the better informed firm to worse informed market participants. However, that alone is not enough. Think of (i) a management which exactly knows its firm’s future economic performance as measured by the ability to pay dividends to shareholders, and (ii) capital market participants who have no such information. Even if managers decide to "truthfully" disclose all their information in the form of some earnings figure, information asymmetry prevails, because value is not well-defined on incomplete and imperfect markets. Therefore the properties of the reported earnings figure are unknown. This means that capital market participants do not know about the preferences and information of managers and what expectations about their own and other market participants’ preferences were used to calculate the earnings figure. On the other hand, managers are unaware about what investors are exactly interested in and how they value certain types of claims. The disagreement about value is an integral part of incomplete and imperfect markets and cannot be resolved by accounting, because whenever value is not well-defined, accounting is neither. If markets were indeed perfect and complete, then any accounting information would be redundant, since there would be no asymmetric information but unanimity about the value of claims.
The above discussion makes clear that the decision-useful role of financial accounting consists of two elements. First, financial accounting is about defining a representation of value based on the subjective expectations of regulators or market participants about the set of other market participants, their preferences and their information endowment. In other words, it should approximate the market value of claims on perfect and complete markets under the conditions of imperfect and incomplete markets. This "definition-aspect" of financial accounting is the set of rules on how to calculate accounting figures, such as earnings or equity. Second, financial accounting should report these value representations to worse informed capital market participants. This "reporting-aspect" of financial accounting is the set of rules on how to structure and disclose information.

Financial accounting only works the way it is designed to work if the accounting information sent by firms is accurate. Financial accounting accuracy means that both the firm insider and the capital market participants know and obey the rules of producing and transmitting information about states. Financial accounting accuracy must not be confused with financial reporting quality. Accuracy means that the signals sent are related to the states in accordance with the systematic linkage that has been previously defined. Quality however refers to the structure of the links between signals and states, regardless of what type of signal is sent. In other words, financial accounting quality only means that accounting signals allow recipients to narrow down the expectations about states of the world, whereas financial accounting accuracy additionally requires that exactly the signals intended by the rules are used. To give an example, if managers overstate earnings, but investors exactly know the managers’ incentives and thus are able to fully subtract the overstatement, the reported earnings are inaccurate but have the same quality as accurately reported earnings. However, in general capital market participants do not know

---

2 Bluntly said, an accounting system that does a perfect job in reducing information asymmetry makes itself obsolete.
the manager’s incentives and cannot correctly infer overstatements. Therefore inaccurate earnings in most cases have a lower quality than accurate earnings.

There are different reasons why accounting information is inaccurate and therefore not credible. First, the firm insiders who produce the financial reports (managers or employees working in the accounting department) might be to some degree uninformed about past, current or future states of nature. This risk increases in the complexity of the firm’s business and decreases in the firm insiders’ expertise. Second, despite rapid advancements in data-collection and -processing, financial reports are ultimately produced by humans who make errors. Finally, there are a conflicts of interest between firm insiders and external market participants, that is insiders have incentives to utilize their information advantage at the outsiders’ expense by intentionally misrepresenting the economic reality through financial reports.

To avoid that financial reports become inaccurate, numerous institutions are situated alongside the road from raw economic transactions to the published financial report about those transactions. The most important ones are the following: First, the shareholders should install a competent management, which in turn should hire, train and lead a competent accounting department staff. Second, it is the board’s responsibility to monitor the managers and their accounting practices, thereby making sure that they act and report in the shareholder’s (and other external stakeholders’) interest. Third, the firm’s internal control system (ICS) ongoingly checks whether transactions are recorded in ac-

---

3 The distinction between accuracy and quality is particularly important when it comes to analyzing effects of accounting system characteristics, such as conservatism or the tightness of standards. The intriguing question then is how the accounting system characteristics jointly change the signals sent by the firm (accuracy) and the signal-state relations on the side of financial statement readers (quality). Examples for this type of research include the papers by [Ewert / Wagenhofer (2005)] and [Ewert / Wagenhofer (2016b)].

4 It might well be that inaccurate earnings have a higher quality than accurate earnings. This is because quality in the sense of fineness is a natural ranking criterion for information systems, whereas the prevailing accounting rules only exist by the definition of regulators or market participants. If the “market for accounting rules” is imperfect and incomplete (which it obviously is), the prevailing accounting systems are not the ones that have the highest quality in relation to what they cost the economy.
cordance with *Generally Accepted Accounting Principles* (GAAP) and result in a "true and fair" presentation of the firm’s economic stance. Fourth, an independent external auditor performs a periodical statutory audit of the firm’s separate and consolidated financial statements. Moreover, the auditor serves as an independent financial accounting expert at certain occasions that require a professional assessment of financial reports and accounting-related practices (valuation opinion, due-diligence audits, voluntary audits). Fifth, after being published, financial reports are at random checked for their accuracy and GAAP-conformity by enforcement institutions. Enforcement institutions either initiate a lawsuit or "name-and-shame" misreporting firms. Both consequences are costly for the firm ex post, and thus provide incentives to invest in accurate reporting ex ante. Sixth and finally, market participants themselves can sue the firm or its management, in case financial reports turn out to be wrongful and they feel deprived of their financial claims.

This dissertation deals with the contribution of the external audit to the informative value of financial reports. Compared to the numerous institutions that exist to render financial reports accurate and credible, the contribution of the external audit to the value of financial reporting is particularly large. The external auditor is a proven expert in the field of financial accounting. As business transactions can be highly complex, so are the accounting rules. This gives a lot of room for both unintentional errors and intentional misstatements by the auditee. The external auditor can detect such misrepresentations better than anybody else. Moreover, the auditor makes the improvement in financial reporting accuracy accessible to outsiders who are non-experts in financial accounting through his or her audit opinion. However, as any other seal of quality, the external audit can only add credibility to financial statements if auditors are in fact and are perceived by outsiders as (i) competent, (ii) diligent, and (iii) independent from the audited firm at any stage of the audit. Even though the auditing profession has an excellent reputation and the selection criteria
for auditors are rigorous, spectacular accounting scandals in which auditors played key roles showed that high quality audits cannot be taken for granted. First, auditors are human, rational utility-maximizing decision-makers. As such, they are averse to exerting costly work effort. Given that their actual effort during the audit is never observable, they have an incentive to shirk and blame bad luck for a wrong audit opinion. The traditional way to overcome the moral hazard problem in principal-agent-relationships, which is to let the agent participate in the outcome of his or her productive effort \[ \text{[5]} \] does not work, because auditors would then have incentives to collude with the management. Making the auditor’s compensation contingent on the audit opinion is also not a feasible solution, since auditors could report whatever opinion just to maximize their compensation, regardless of the evidence they did (or did not) collect. Second, auditors are entrepreneurs, who are hired by the firms to be audited. Therefore an auditor is economically dependent on the auditee. This dependence creates incentives to consciously distort the audit opinion in the auditee’s favor to secure the audit engagement for the future.

This dissertation addresses the general question of how to optimize (not necessarily maximize) external auditors’ incentives to provide high quality audits, in order to optimize the contribution of external audits to the informative value of financial reports. Audit quality is hereafter defined as the probability that material misstatements in financial reports are detected by the auditor before they are publicly disclosed, and those reports are either marked as ”materially misstated” by a qualified audit opinion, or the correction of the material misstatements is initiated by the auditor. In order to supply high quality audits, auditors have to be (i) diligent in conducting the audit, and (ii) independent from the auditee in reporting the outcome to external stakeholders of the auditee. Parallel to a lively ongoing regulatory debate, two streams of auditing research have developed around these issues. This dissertation deals with the first issue, that is the incentives of the auditor to exert high effort during the audit. It does not address questions related to

\[ \text{[5]} \] See the seminal work by Holmström (1979).
INTRODUCTION

Figure 1.1: Financial accounting and auditing in a capital market context

the issue of the auditor’s independence from the client.

Figure 1.1 pictures the role of financial accounting and external auditing in a capital market context. It is a graphical representation of the stylized economy and its interaction interfaces that are analyzed in the course of the dissertation.

1.2 State of Research

Auditing research, and in particular theoretical auditing research originates from a handful of seminal papers in the late 1970s and 1980s. These papers show how auditing can generate economic value (see Townsend (1979) and Magee (1980)), how auditors can be incentivized to exert audit effort (see Antle (1982), Fellingham / Newman (1985) and Nelson / Ronen / White (1988)) and how the fact that auditors are economically dependent on their clients may threaten the independence of their judgment (see DeAngelo (1981a) and DeAngelo (1981b)).
INTRODUCTION

Given that auditors are rational utility-maximizing decision-makers, they can be incentivi-
zed to exert effort either by gains or losses in their present and future consumption
potential. Because outcome-contingent audit fees are prohibited, these gains and losses
cannot be wrapped up in monetary payments from the auditee to the auditor. The first
of the main auditor incentives is reputation. In an information-economic sense, a gain
(loss) of an auditor’s reputation is an increase (decrease) in the subjective belief of current
and potential future auditees or capital market participants that the auditor is of some
desirable good type (e.g. industry expertise). An auditor with a higher reputation is
preferably hired by firms. Therefore, those who are of the good, desirable type have an
incentive to separate from the bad, undesirable type by doing what the bad type cannot
do (e.g. provide industry-specific, high quality audits). The effects of auditor reputation
are scarcely analyzed in the literature. Notable papers that focus on how reputation in-
centives (endogenously) emerge and affect the auditor’s choice are Datar / Alles (1999),
Bigus (2006), Corona / Randhawa (2010), and Bigus (2011). This dissertation does not
deal with reputation incentives.

The second main auditor incentive is legal liability. If an auditor fails to detect material
misstatements in financial reports and the auditee and/or investors suffer losses due to
the publication or reliance on the misstated report, the auditor can be sued for a breach of
his or her duty to protect capital market participants from wrongful financial information.
Hereafter this situation is called an ”audit failure”. If an impartial court concludes that
the auditor has been negligent in conducting the audit, he or she has to compensate
the plaintiffs for their losses. In addition courts may also award punitive damages to a
negligent auditor. The threat of suffering from legal liability provides a strong incentive
to invest effort into a diligent audit ex ante. In most cases an audit failure is based on
the non-detection of a material overstatement (type II error). The incorrect classification
of an accurate report as materially overstated is deemed less severe. On the one hand,
foregoing profits due to an overly pessimistic audit opinion is less harmful than making
bad investment due to an overly optimistic audit opinion. On the other hand, the client has a chance to overrule the wrong audit opinion by insisting on the favorable report if he or she is convinced that the auditor is wrong.

Auditors’ legal liability towards third parties is by far the most-investigated incentive in theoretical auditing research. The reason for the strong research interest is that legal liability regimes are designed by law- and policymakers, and research can give immediate guidance on how the rules should or should not look like. The first wave of influential theoretical papers on auditors’ legal liability has been published in the top-tier U.S. accounting journals during the 1990s. The Anglo-American legal system is and has always been highly litigious compared to Middle-European legal systems.\(^6\) However, even by the harsh U.S. liability standards, auditors’ third party liability has been perceived as being excessive in the 1980s and early 1990s. In particular, (i) the joint-and-several liability of auditor and auditee, and (ii) the extension of the scope of third party liability from third parties who are in a "quasi-privity of contract" towards the indefinite group of capital market participants whose use of the audited reports the auditor "could have had reasonably foreseen" became a heavy burden for the auditing profession. In reaction, the Securities and Exchange Commission (SEC) issued the Private Securities Litigation Reform Act of 1995 (PSLRA), which notably relaxed auditors’ third party liability in various aspects. The second wave of research on the enforcement of audit quality through legal liability emerged in the aftermath of the spectacular accounting scandals of Enron and WorldCom around the millennium. The collapse of the leading accounting firm network Arthur Andersen LLP, which played a key role in the fraudulent accounting practices of Enron and WorldCom, forced the regulators to act. To restore the confidence of capital markets in external audits and to prevent a contagion of the whole audit industry, the SEC issued the Sarbanes-Oxley Act of 2002 (SOX). One main feature of the SOX was the creation of the Public Company Accounting Oversight Board (PCAOB) as an external

\(^6\) This is most prominently documented in the study by La Porta / Lopez-de Silanes / Shleifer (2006).
oversight and enforcement authority for the formerly self-regulated audit profession.

Existing theoretical research on the enforcement of audit quality through legal liability can be classified into four categories alongside the attribute dimensions of liability systems. The first important attribute of liability is the absolute extent of (financial) consequences for the defendant. The respective literature sub-strand asks how extending or limiting the amount of legal liability (e.g. through liability insurance or incorporation) auditors are exposed to affects their behavior. The most important papers addressing that question are Nelson / Ronen / White (1988), Dye (1993) and Dye (1995). The papers by Dye (1993) and Dye (1995) are considered as path-breaking in the modeling of auditor incentives and the market for audit services. Other important papers include Moore / Scott (1989), Melumad / Thoman (1990), Schwartz (1997), Ewert / Feess / Nell (2000), Newman / Patterson / Smith (2005), Laux / Newman (2010), and Deng / Melumad / Shibano (2012).

The second sub-stream of theoretical research on auditor liability investigates under which conditions auditors should be held liable. The most frequently asked question is whether auditors should be considered negligent whenever an audit failure occurs (strict liability), or only if they verifiably violated pre-defined standards regarding the quality of audit procedures (negligence liability). Notable papers include Balachandran / Nagarajan (1987), Dye (1993), Narayanan (1994), Schwartz (1998), Radhakrishnan (1999), and Liu / Wang (2006). Given a negligence liability system, the negligence standard can either be precisely defined or ”vague”. A standard is said to be ”vague” if the auditor has a basic expectation, but does not exactly know whether a court will consider his or her audit procedures as diligent or negligent in case of an audit failure. Intuitively, one might expect that legal uncertainty is costly, because it sometimes leads to under- and over-investments into audit effort. However, Willekens / Steele / Miltz (1996), and in particular Ewert (1999) show that the imprecision of the audit standard exhibits an additional effort-incentive, because the auditor has the chance to escape liability also in
INTRODUCTION

case of an audit failure. Finally, even if an auditor is found negligent, he or she is not
necessarily held liable, because he or she might agree with the plaintiffs on an out-of-court
settlement. The incentive effects of such settlements are discussed by Zhang / Thoman
(1999).

The third literature sub-stream deals with the question how the damages should be ap-
portioned between the auditor and the client. The contributions are centered around the
establishment of the PSLRA in 1995 which replaced the rigorous joint-and-several liability
by a more lax variant of the proportionate liability system. Consequently the focus of the
papers, which include Narayanan (1994), Chan / Pae (1998) and Hillegeist (1999), lies
on a system comparison. However, the difference between the systems is only relevant if
some of the defendants are bankrupt. Then the solvent defendants are jointly liable for the
bankrupt defendants’ damages under the joint-and-several liability regime, whereas they
are only liable for their own shares of damages under the proportionate liability regime.
If all defendants remain solvent, then the damages are shared proportionately according
to the court’s assignment of negligence. The question how the damage apportionment
affects the auditor’s and the auditee’s incentives to invest into accurate reporting (or mis-
reporting) in the latter case is under-addressed in the literature, with Hillegeist (1999)
being the only exemption. Hillegeist (1999) shows that a more auditor-weighted sharing
of liability makes the auditor work harder, but also increases the amount of misreporting
by managers. The latter effect dominates, which is why the accuracy of financial reports
decreases as auditor liability gets tightened.

The fourth and final attribute dimension of auditors’ legal liability is the scope of liability,
that is the group of market participants the auditor can be held liable towards. In
general, auditors can be held liable towards the auditee for a breach of the audit contract.
However, there is no contractual relationship between the auditor and outside capital
market participants. These ”third parties” cannot sue the auditor for a breach of contract,
but for a breach of his or her duty to protect them from wrongful financial accounting
information. The crucial question is, to whom the auditor has this duty. The only notable paper addressing this important question is Chan / Wong (2002). Chan / Wong (2002) compare the three legal interpretations regarding the scope of auditors’ third party liability that historically evolved in the U.S. Under the narrow Ultramares-approach the auditor can only be held liable for the losses of third parties who are in a quasi-privity of contract with the auditor, because their identity and their use of the audit opinion is known to him or her. Under the Restatement-approach auditors can also be held liable for the losses of third parties who the auditor does not know in person, but whose use of the audit opinion is known to him or her. Finally, under the liberal foreseeability approach the auditor can be held liable also for the losses of those market participants whose use of the audit opinion he or she could have had reasonably foreseen. In the extreme, this refers to each and every investor on the capital market. Somehow counter-intuitive, Chan / Wong (2002) conclude that the Ultramares-approach is superior because it provides the least investor protection. This makes debt financing rather expensive, thereby ”self-restraining” the firm insiders’ incentives to obtain additional cheap debt and shirk.

Although these four sub-streams of literature produced a mass of interesting and largely unexpected results over the last decades, our understanding of the auditor’s role in a capital market context is still rather limited. In particular, two problem fields can be identified. First and most important, existing research has a narrow focus on audit quality. In many papers, the auditor is viewed as the monopolist supplier of financial information. The firm is either unable to actively improve the accuracy of preliminary reports (see Laux

---

7 This legal interpretation is named after the case Ultramares Corporation v. Touche Niven & Co., 255 N.Y. 170, 174 N.E. 441 (1931). It is also called known user approach.
8 This legal approach is named after the treatise Restatement of Torts (Second) by the American Law Institute, which is its (non-legislative) basis. It is also named foreseen user approach.
9 Under the more liberal approaches additional debt is cheaper because more investors are insured against losses through auditor liability. This enforces the managers’ incentives to obtain additional debt at the cost of old debtholders and shirk. In order to commit to productive work when raising the original debt, the managers have to strongly restrict the investment amount. This under-investment problem is socially detrimental, and less severe under the narrow Ultramares-approach; see Chan / Wong (2002), pp 116 et seq.
INTRODUCTION

Newman (2010) for a typical example), or even plays against the auditor by engaging in active misreporting (see Hillegeist (1999) for a typical example). Of course, both approaches have their merits, because they blank out aspects of reality in order to identify the focal economic effects in better contour. For example, the model of Laux / Newman (2010) allows to work out the incentive effects of various forms of regulation on the auditor’s effort choices. The setup of Hillegeist (1999) on the other hand, gives insights on how legal liability incentives facilitate or impede the auditor’s combat against earnings management and accounting fraud. However, what seems forgotten is that auditing is not the sole source of accurate financial information, but only a single, rather downstream link in the chain from raw, partly erroneous transaction records to accurate, GAAP-conform financial statements. I name this first problem field the ”multiple information-source-complex”. Notable exceptions are the papers by Pae / Yoo (2001), Patterson / Smith (2007), Ewert / Wagenhofer (2016a), and Ewert / Wagenhofer (2018).

The second problem of the literature is that the drivers of audit quality are almost always analyzed in isolation. Typically, papers study the changes of equilibrium behavior after a variation in one attribute of the legal liability system, while blanking out all the other attributes. For example, Liu / Wang (2006) study the transition from a strict liability regime to a vague negligence liability regime, assuming (i) a constant liability in the amount of the investors’ loss (”out-of-pocket”-liability), (ii) a fixed group of investors with legally justified liability claims, and (iii) bankruptcy of the reporting firm, which makes the auditor the sole defendant. Again, this is a good approach to work out the incentives provided by distinct institutional features. However, legal liability systems are a complex and entangled mesh of attributes. The interaction effects between the different attributes, that is the incentives that are jointly created by different characteristics of liability systems are yet rather unexplored. Among the few exceptions are the papers by Balachandran

---

10 In Laux / Newman (2010) not only the choice of audit quality, but primarily the auditors client acceptance decision is analyzed.
INTRODUCTION

Figure 1.2: State of research on the auditor’s contribution to accurate financial reporting

Nagarajan (1987) and Ewert / Feess / Nell (2000), who jointly study negligence rules and a limitation of the absolute liability amount through insurance. Moreover, not only the interaction effects within the liability system, but also the interaction of legal liability with other legal institutions, such as board oversight or enforcement authorities are rather unknown. A very recent paper that stands as an example for what could be done is Ewert / Wagenhofer (2018). Finally, the interaction between legal liability and auditor characteristics, such as wealth or risk attitude are almost completely untouched. The only notable exception is the seminal paper by Dye (1993) who studies the interaction of the negligence standard of the liability system, and the auditor’s wealth. I label this second problem field the ”liability-interaction-complex”.

In summary, figure 1.2 pictures the current state of research on the drivers of audit quality, respectively its contribution to financial reporting accuracy. The arrows thereby represent interaction effects on which distinct research has been done yet.
1.3 Research Questions and Approach

1.3.1 Main Research Question

This dissertation strives to give an answer to the following general research question:

*How does the enforcement of accurate financial reporting through third-party liability affect the auditor’s contribution to the detection of errors in financial statements prior to their publication?*

This formulation already puts a few boundaries on the scope of the dissertation. First, I am interested in accurate financial reporting, that is the question whether financial statements of the firm represent objective economic facts in accordance with generally accepted accounting rules and principles. I abstract from questions regarding financial reporting quality by assuming that if it is common knowledge that the firm follows the accounting rules and principles, the information asymmetry about the prevailing economic state is perfectly resolved. In other words, if the firm’s financial statements are accurate, then they also have the highest attainable quality. Second, the inaccuracies that are to be avoided by the audit are unintentional errors, that is misrepresentations resulting from (i) fundamental economic uncertainty and (ii) uncertainty about the accounting treatment of transactions. That said, the dissertation does not explicitly deal with intentional accounting distortions, such as window dressing of financial statements or even accounting manipulation. From an abstract point of view, one could also assume that the misstatements are intentional, because refraining to invest into the prevention of extant errors in order to save costs is qualitatively equivalent to seed a misstatement into financial statements in order to gain a financial advantage. However, at a more practical level, intentional distortions of financial statements exhibit different economic mechanics than negligence regarding corporate governance activities. Third, I focus on the auditor’s choice of productive effort during the audit process, but not on his or her reporting choice after having accumulated and evaluated the evidence. This means, I restrict attention
to unintentional misrepresentations also at the audit stage, but abstract from reporting issues such as auditor conservatism or collusion with the client. Overall, the dissertation therefore investigates the joint production of information by the auditor and the client, whilst not addressing questions regarding the redistribution of existent information.

At this point, one might argue that studying the production of new information is not helpful to understand situations of asymmetric information, in which by definition some information already exists. However, information asymmetry can not only be resolved by transmitting the existent information from the better informed insider to the worse informed outsider. It is often the case that information cannot, or only at very high costs can be credibly communicated from the insider to the outsider. Then it is necessary or preferable for the outsider to produce the information on his or her own, respectively initiate the production of information by an information intermediary such as the auditor. To study this production process it is helpful to completely abstract from reporting issues by assuming symmetric information among those who are involved in the information production process. Hence, whenever the reader is confronted with the (technical) assumption of symmetric information distribution hereafter, he or she should bear in mind that there is an underlying information asymmetry that the activities of the firm and the auditor aim to reduce. This underlying information asymmetry prevails at different levels. First, the firm possesses more information about its transactions and economic facts than the auditor and external stakeholders at the beginning of the audit. However, for some reason the firm is unable to cast this information into GAAP-conform accounting figures. Therefore, with regard to the correct translation of information into accounting figures that are universally understood and thus communicable, both the firm and the auditor are initially ignorant. Second, information is asymmetrically distributed within the firm. For example, the production division might report about the profitable prospects of a planned investment project. However, without further information (e.g. information about the product managers’ reporting incentives), the accounting and control departments, as well
as the external auditor and outside stakeholders of the firm cannot assess the veracity of that information and are thus initially equally uninformed about the true prospects of the project. Third and most basic, the firm might in fact be as uninformed as the auditor and its external stakeholders. In this case the better informed inside party is some person or institution outside the reporting firm, or simply nature.

As it is obvious from figure 1.2, the large part of the complex effect network around the main research question is yet unexplored. It is needless to say that addressing the main question from the view of each and every institutional interface pictured in figure 1.2 goes beyond the scope of any scientific piece of work. Therefore I focus on a few institutional complexes and issues which have received particularly little attention by extant research. The aim of this dissertation is to significantly densify our understanding about the drivers and barriers for audit quality, while also pointing out some avenues for future research. I want to emphasize that I do not attempt, or even claim to ”close research gaps”, which is a catchphrase frequently used in scientific jargon. The definitive closing of a research gap requires to reduce the phenomena of interest to the underlying axioms of reality. However, the research objects of the social sciences are vibrant, constantly changing preferences, relationships and interactions of individuals. Based on the present state of knowledge there are hardly sufficient axioms to which these social fabrics can be reduced.\footnote{11} Moreover, as the overwhelming majority of economic studies, this dissertation consists of narrowly bounded partial considerations, which blank out a huge part of reality for the sake of logical analyzability. Even if sufficient axioms existed to explain the phenomena of interest, the contribution of such partial considerations is by definition never sufficient to close a research gap.

\footnote{11 To give an example, the assumption that individuals \textit{strive} to maximize their individual utility can be considered as an axiom of neoclassical microeconomics. However, the assumption that individuals \textit{do} always choose the action that maximizes their individual utility is not an axiom, as there is plenty of evidence for the bounded rationality of decision makers. Yet, the assumption is reasonable, and in many situations descriptive.}
1.3.2 Dissertation Papers

The dissertation consists of three separate research papers, each of them pursuing the main research question from a different angle. Each paper is prepared with the intention of being submitted to an international, peer-reviewed journal. In the following, I briefly sketch the research approach and questions of each paper. However, I leave the detailed outlines of the contributions, as well as further clarifications of the research questions to the introductory sections of the respective papers.

The first paper is entitled "The Impact of Damage Apportionment on ICS Quality and Financial Reporting Accuracy". This paper asks the following main questions:

- How does the apportionment of legal liability between the auditor and the auditee in case of an audit failure affect the quality of internal controls, the quality of audits, and the accuracy of published financial reports?

- How is the efficiency of the resource allocation between the auditor and the auditee affected by the damage apportionment, and how should a "socially optimal" damage apportionment rule look like?

This paper tackles both the multiple information-source- and the liability-interaction-complex. First and more obvious, the paper accounts for the fact that financial reporting accuracy is the joint production outcome of the auditee and the auditor. The auditee’s activity of interest is the investment into the internal control system. As the auditor and the auditee share the production of accurate financial reports, material misstatements are both attributable to negligent auditing and negligence regarding internal controls. Consequently, both of them can be held liable for the damages of investors who relied on misstated financial reports. The fact that the auditor and the auditee each make their individual contribution, but their respective liability consequences are based on the joint outcome creates a strategic interdependence of efforts. This means that the
INTRODUCTION

auditors' choice of audit quality depends on the audited firm's investment into the ICS et vice versa. Second, the paper also investigates how multiple legal and environmental factors interact and jointly affect decisions about the investment in financial reporting accuracy. Specifically, I study the joint effects of (i) the damage apportionment within a proportionate liability, and (ii) the scope of legal liability. This is an interesting question, because there is a common perception that auditors have a responsibility towards the investing public that transcends contractual and quasi-contractual relationships. This means that they have to be aware that capital market participants, whose use of the financial report is not known to them, will rely on their professional opinion. The stronger the baseline tendency to hold auditors liable is, the more investors will attempt to, and eventually succeed in recovering their losses by suing the auditor.

The second dissertation paper is entitled "Audit Quality, Legal Liability, and the Audit Market under Risk-Aversion". This paper deals with the following main questions:

- How do legal liability characteristics, such as the absolute amount of liability, and the prevailing negligence regime affect the choice of audit quality, given that auditors are risk-averse?

- How do these legal liability characteristics affect the audit market structure?

This paper is exclusively dedicated to the liability-interaction-complex. The starting point of the paper is the ubiquitous simplifying assumption of auditors being risk-neutral decision-makers. However, there is overwhelming evidence that in reality individuals are risk-averse. In this paper I study the yet unexplored interaction between the design of liability systems and the degree of auditors' risk aversion. In a first step I analyze each individual auditor's decision about the supplied audit quality in isolation. I then integrate this decision into a market equilibrium between the supply and demand for audit services. This approach accounts for the fact that it does not matter what audit
quality any arbitrary auditor would choose, but what audit quality those auditors who participate in the market because they are actually hired, supply. Finally, I investigate how auditors’ legal liability shapes the audit market structure. Upholding audit quality in the mid- and long run requires the audit market to be stable in the sense that the supply and demand for audit services are diversified across a large number of auditors, instead of a few global networks (”Big-N”) dominating the market and being ”too big to fail”. The accounting scandals of Enron and the subsequent disintegration of the global audit network Arthur Andersen in 2001 gives an idea what kind of havoc the collapse of another large audit firm network would cause. This paper puts special emphasis on giving recommendations to regulators and policymakers on how to foster the audit quality of the currently dominant large firms, while also improving the competitiveness of smaller firms in order to de-concentrate the industry.

The third paper of the dissertation is entitled "Sequential Auditor-Client Interaction under Strategic Effort Complementarity". This paper is centered around the following question:

- How is the allocation between firm-side corporate governance activities and external audit effort affected by strategic substitution or strategic complementarity between the respective efforts?

- How do changes in the amount of legal liability affect this allocation and the aggregate accuracy of financial reports?

- How do differences in stakeholders’ demand for firm-side governance effort and audit quality affect the resource allocation?

This paper refers to the multiple information-source complex and focuses on the "production technology" used by the firm and the auditor to generate accurate financial statements. The starting point of this paper is a discrepancy between theoretical predictions
and empirical evidence on whether firm-side governance activities, such as internal controls or audit committee work, and the external audit are substitutes or complements. The theoretical literature backs the substitution-view by arguing that a better corporate governance reduces the risk of material misstatements in financial reports, thereby allowing the auditor to save on costly audit procedures. Empirical studies finding a complementary relation conversely argue that the increased investment into one corporate governance instrument (e.g. internal controls) increases the demand for other corporate governance instruments, such as the external audit. I study how the strategic relation within the common production technology, that is whether the firm side activities increase or decrease the marginal effectiveness of auditing, affect the total effectiveness of the audit. I then analyze how exogenous shocks, such as changes in the auditor’s legal liability exposure, affect the strategic relation and the observable resource allocation. Moreover, I study how differing stakeholder demands, for example the owners’ demand for internal controls to measure managers’ performance, and the investors’ demand for audit quality to avoid bad investment, affect the resource allocation. Finally, I subject the theory to empirical scrutiny by running a cross-country regression analysis.

The following figure 1.3 illustrates the contribution of the dissertation to the research field. The bold arrows indicate areas in which the dissertation deepens existing research. The dotted arrows indicate aspects and interdependencies dealt with in the dissertation, which are yet uncharted scientific territory.

1.4 Methodology

This dissertation strives to identify new, yet unexplored causal effects in the triangular auditor-firm-investor-relationship. The identification of a causal effect-chain requires a theoretical foundation. This means that an empirical phenomenon or conventional wisdom is cast into a logical framework, and then reduced to basic and accepted assumptions
about the reality we live in. This integration of new elements into a logically consistent and coherent framework is what is understood as "building a theory". It is part of human nature that the curiosity for new theories is fueled by the unexpected and uncomprehended. Hence, theoretical research often produces findings that are surprising and go against "reasonable conjectures". In this regard, the dissertation makes no exception.

In all three dissertation papers I use mathematical instruments in the form of decision-and game-theoretic modeling to develop theories that propose reasonable answers to the outlined research questions. The three papers proceed in the same way. First, the model is set up, which means that the relevant aspects of the assumed state of the world are formally described. Then, a stable equilibrium, that is a situation in which none of the parties changes his or her behavior given the other parties’ behavior, is established. The centerpiece of each analysis is the subsequent comparative static analysis, that is an analysis of the change in the equilibrium behavior with respect to a change in one or more of the exogenous model-parameters. The third paper "Sequential Auditor-Client Interaction under Strategic Effort Complementarity" has a special status, because it not only lays out a theory, but also puts this theory to the data.
2 The Impact of Damage Apportionment on ICS Quality and Financial Reporting Accuracy

ABSTRACT: This paper investigates how the damage apportionment between the auditor and the auditee affects (i) the quality of the internal control system, (ii) the supplied audit quality, and (iii) social welfare. The analysis takes place in a setting, in which the auditor is not only liable towards the primary addressees of the audited report, but assumes a public responsibility towards the capital market. Using a game-theoretic model, I show that shifting liability away from the auditor, towards the client can lead to higher audit quality but lower ICS quality. I provide a new argument why internal controls and external auditing can be either complements or substitutes, as it is suggested by the mixed empirical evidence. Furthermore, assigning all the damages to the auditor leads to a Pareto-efficient resource allocation between ICS quality and audit effort. This finding justifies the common 'deep pockets'-assumption in the auditing literature.

I thank Derek Chan (Discussant EIASM Workshop), Travis Dyer, Ralf Ewert, Heinz Königsmaier, Sebastian Kronenberger, Eva Labro, Wayne Landsman, Dennis Voeller (Discussant VHB), Alfred Wagenhofer, Stefan Wielenberg, session participants at the 39th Annual Congress of the EAA in Maastricht, 2016, session participants at the 78th VHB Annual Meeting in Munich, 2016, session participants at the 12th EIASM Workshop on Accounting and Economics in Tilburg, 2016, participants of the workshop at the University of North Carolina at Chapel Hill, participants of the 5. iwp-Wissenschaftsforum in Vienna, 2018, and my fellow PhD students Robert Eder, Harald Hinterecker, Claudia Langbauer, and Felix Niggemann for their valuable comments which helped to significantly improve the paper.
2.1 Introduction

This paper studies how the apportionment of legal liability between a firm and its external auditor affects the effectiveness of the firm’s internal control system, the supplied audit quality, the overall accuracy of published reports, and social welfare. The reporting firm’s ICS and the external audit are both main pillars of accurate financial reporting. The importance of the ICS has been strongly underlined in the U.S. by the notorious Section 404 of the SOX. Section 404 raised the requirements for internal controls drastically and legally obliges listed companies to install, maintain and document an adequate ICS. In case the audited reports ex post turn out to be materially misstated and share- or debtholders suffer damages, these damages are not only attributable to insufficient auditing, but also to the firm’s negligence regarding internal controls. Investors will hence attempt to recover lost capital from both the firm they misinvested in, and the auditor who issued the wrong opinion. However, the auditor’s liability does not end with the apportionment of a damage share. The prevailing legal view is that auditors fulfill a legal mandate in investor protection and serve a public interest. In the 1984 case United States v. Arthur Young & Co., the U.S. Supreme Court stated:

By certifying the reports that collectively depict a corporation's financial status, the independent auditor assumes a public responsibility transcending any employment relationship with the client. The independent public accountant performing this special function owes ultimate allegiance to the corporation’s creditors and stockholders, as well as to the investing public. This 'public watchdog’ function demands that the accountant maintain total independence from the client at all times and requires complete fidelity to the public trust.

This judicial viewpoint leaves room for a broad legal interpretation of the auditor’s liability

---

12 The term ICS hereafter refers to accounting-related internal controls only.
scope. Taking the 'public watchdog' argument literally, an auditor can be held liable not only for the losses of known users of the audit opinion, but also for losses of market participants, who have based unfavorable decisions on a wrong audit opinion and whose use of the audit opinion the auditor could have had reasonably foreseen. How many of those 'foreseeable third parties' jump on the bandwagon of auditor liability, and how successful their lawsuits are, depends on the basic tendency to hold auditors liable for investor losses.

The main contribution of the analysis is that it shows when and how the damage apportionment between an external auditor and the client firm actually matters. Proportionate liability is the prevalent system across the world. Unfortunately, existent analytical research gives no guidance on how the damage apportionment within this system economically works and how it should ideally look like. In the standard Dye (1993)-type auditor-client-investor game splitting the damages between the auditor and the client does not affect the equilibrium, because the firm bears the auditor's liability either through the audit fee or through a reduction in the firm's selling price anyways. However, the damage apportionment becomes important as soon as there is a break in the auditor's and the client's liability regimes. Then the damage apportionment determines how much incentives of the auditor's liability regime spill back into the firm's decision et vice versa.

In this paper, the difference in the liability regimes stems from auditors' status as the financial markets' gatekeepers. Because of that special role their liability is not confined to the immediate capital providers of the firm, but can extend to whomsoever the courts consider as foreseeable third parties.

I use a game-theoretic model, in which an investing firm and its auditor sequentially interact. First, the firm spends resources on the installation of an ICS. Subsequently the auditor audits the firm's financial statements to eliminate remaining errors. If neither the ICS nor the auditor detect an overstatement, outside investors finance an investment project of the firm. In addition, a group of investors invests capital into projects that
are economically related to the audited project. If the published statements turn out to be materially overstated, the project investment is lost and both the client and the auditor are held liable according to a pre-defined damage apportionment rule. The auditor additionally has to compensate foreseeable third parties, that is the investors of the related projects, for their losses.

The analysis yields the following insights: First, shifting liability away from the auditor towards the client firm can deteriorate the quality of the firm’s ICS. The reason for this seemingly unintuitive finding is that any variation in the damage apportionment varies the amount of the auditor’s additional liability towards foreseeable investors. As a consequence, the client decides between (i) improving the financial reporting accuracy on his own by investing in the ICS, or (ii) inciting the auditor to exert disproportionately more effort by lowering the ICS quality first. Which strategy is favored in equilibrium depends on the size of the auditor’s damage share. Second, in case (ii) the relaxation of auditors’ liability is followed by an increase in audit quality. Third, shifting liability away from the auditor strictly decreases the accuracy of audited financial reports when the damage apportionment is currently sufficiently even, because both ICS and audit quality decline. Fourth, if the auditor bears all the liability, the equilibrium that evolves out of the game is Pareto-efficient, because the liability incentives of both defendants are aligned through the audit fee. For all other damage apportionment rules, an efficient equilibrium is not attainable. However, in general there exists a game characterized by a non-zero damage share assigned to the client firm which yields an inefficient equilibrium that is superior to the attainable efficient equilibrium.

These results contribute to the literature and practice in several ways. First, and most important the paper stresses that the joint production of information by the auditor and the client requires a joint analysis of liability as well. Second, the analysis shows that the optimal damage apportionment depends on the cost-efficiency of the audit-technology, which is particularly relevant for regulators in the light of innovations like big-data, blockchain
and artificial intelligence. Third, the model can explain the mixed empirical evidence on whether external auditing and internal controls are substitutes or complements. Finally, the finding that assigning all the damages to the auditor elicits an efficient equilibrium justifies the common model assumption that the client is bankrupt and damages can only be recovered from the auditor.

2.2 Related Literature

This paper complements on three strands of theoretical auditing research. First and most important, it adds to the literature on the economics of damage apportionment between the auditor and her client. Second, it complements on existing research regarding the interplay of auditing and internal controls. Third, it adds to the scarce literature on the scope of auditors’ third party liability.

The most relevant papers on damage apportionment rules are the ones by Narayanan (1994), Chan / Pae (1998) and Hillegeist (1999), which have all been published around the time of the Privates Securities Litigation Reform Act of 1995. Narayanan (1994) and Chan / Pae (1998) compare the mild proportionate liability system with the strict joint-and-several liability system. However, the difference between the two liability systems is only relevant if the client firm is bankrupt after an audit failure. In my model the client firm always remains solvent, which allows me to analyze the whole range of different damage apportionment rules within the proportionate liability system. To my knowledge, Hillegeist (1999) is the only paper which investigates variations in the effective damage apportionment within the proportionate liability case. He concludes that if the firm strategically reacts to the auditor’s choice of audit quality, a higher damage share of the auditor strictly increases audit quality, but strictly decreases the overall reporting
accuracy due to the high misreporting propensity\textsuperscript{14} \textsuperscript{14}Hillegeist (1999) gets monotone comparative statics for the damage apportionment rule, whereas I find that the firm’s and the auditor’s equilibrium effort levels are (inversely) U-shaped in the damage shares. The reason for this difference is that in Hillegeist (1999) the firm remains solvent only with some probability, whereas the auditor is always solvent. Hence, a liability shift towards the auditor not only incites the auditor to work harder, but also directly increases investor protection. Both effects increase the firm’s expected selling price and make misreporting more attractive to the manager. In my model however, both the auditor and the firm remain solvent with certainty and the auditor bears additional liability towards other capital market participants. Hence, a liability shift towards the auditor not only incites the auditor to work harder but also has a negative effect on the firm because it increases the audit fee (which is equivalent to a lower selling price in Hillegeist (1999)). These countervailing effects are responsible for the U-shapes of the equilibrium curves.

There are also only few analytical contributions that focus on the joint information production of the ICS and the audit. Most closely related to my model are the approaches of Nelson / Ronen / White (1988) and Pae / Yoo (2001). The paper by Nelson / Ronen / White (1988) - among several other aspects of auditing - investigates the interplay of audit effort and the manager’s investment in ICS quality. The authors argue that audit effort and ICS quality substitute each other in equilibrium. Hence, tightening auditor liability simultaneously increases audit effort and decreases ICS quality et vice versa. Overall, the accuracy of audited reports increases in their model if the liability of either the manager’s or the auditor’s liability increases, because the effort increase of the

\textsuperscript{14} See Proposition 4 in Hillegeist (1999), p 364. At first glance the setup of Hillegeist (1999) appears different because he models the misreporting choice of a manager who is informed about the firm state, whereas I model the costly information acquisition of an ignorant firm. However, in both models the firm (respectively its management) trades-off the costs of accurate reporting with the benefit of an increased misreporting probability. In Hillegeist (1999), accurate reporting ’costs’ the management of a low-state firm the private benefit from selling the firm at an inflated price. In my model the benefit of misreporting (implementing a low quality ICS) is the cost advantage of outsourcing accurate financial reporting to the auditor. Hence, in essence the firm-side decisions at the first stage of the game are of the same type.
player whose liability is tightened overcompensates the corresponding effort reduction of the other player. In contrast to Nelson / Ronen / White (1988), not only the two different effort levels but also the liability consequences are substitutes in my model. Hence, I do not consider an absolute liability increase but only a shift of the liability consequences between the players. Furthermore, Nelson / Ronen / White (1988) only consider a static game. Instead, I follow the model structure of Pae / Yoo (2001), who analyze a dynamic game in which the manager chooses his effort before the auditor does. Pae / Yoo (2001) show that - apart from the direct effect of higher auditor liability - in a dynamic setting the anticipation of the auditor’s incentive by the manager creates an additional indirect incentive for the auditor to exert effort. In my model this effect is further enriched by the strategic effects that originate from the damage apportionment. While both Nelson / Ronen / White (1988) and Pae / Yoo (2001) conclude that auditing and internal controls are substitutes in equilibrium, I show that depending on the damage apportionment (i) they might also be complements, and (ii) the substitution might go in either direction.

To my knowledge, the only analytical paper that explicitly deals with the question to whom the auditor should be held liable is Chan / Wong (2002). The authors study how different liability scopes affect a firm’s amount of business investment, the lenders’ willingness to lend sequentially, and the audit effort exerted. In contrary to Chan / Wong (2002), I do not compare different liability regimes by equipping groups of investors with different claims, but vary the amount and success of additional lawsuits against the auditor through the damage apportionment rule. Moreover, I do not concentrate on the investment problem itself. Instead I take the investment as given and look at the firm’s and the auditor’s oversight incentives.
2.3 The Model

The model builds on the basic sequential auditor-client model by Pae / Yoo (2001). It has one period and five stages. The game structure and all parameters are common knowledge. Figure 2.1 sketches the model economy. Figure 2.2 depicts the game tree of the model (decision nodes are grey-shaded).

A risk-neutral firm seeks to extend its business by a new project, which requires funding by external investors of the amount $I$. There are two types of projects on the market. With the prior probability $p$ the project is successful and yields a return with a present value of $Z > I$. However, with the prior probability $(1 - p)$ the project fails. In this case the present value of future returns is zero and the initial investment $I$ is lost. According to the audit risk model the prior probability $(1 - p)$ grasps the firm’s business risk. The investment is a pioneering project within the industry. Its profitability is decisive for the profitability of other investments within the economy. The type of project I have in mind is the Gigafactory 1 project of the electric car manufacturer Tesla. The Gigafactory 1 is a gargantuan battery factory that is currently built in Nevada, U.S. The ambitious goal of
the $5 billion project is to bring down battery costs and pave the way for electric vehicle mass production. The project is very risky, but could be pathbreaking for the whole automotive industry. It follows that the information in Tesla’s audited financial reports is decisive for investments in the development and production of electric vehicles. A priori the investing firm does not exactly know the project type. Hence, the veracity of a good report needs to be checked by the firm’s ICS and a subsequent external audit. Conditional on the outcomes of internal controls and the external audit, the projects undertaken are in expectations successful. Therefore the client prepares a preliminary report, in which he claims to the best of his knowledge that the project is of the good type. This also means that the business risk \((1 - p)\) is fully transferred into the unaudited financial report and is thus equivalent to the notion of inherent risk in the audit risk model.

The probability that an overstated report is detected by the client firm itself depends on the ICS quality \(s \in [0; 1]\). The ICS quality \(s\) is an effort of the client that captures all the activities that aim adapting and improving internal control procedures. Examples for such activities include the hiring and training of employees for the controlling/ internal

\[15\]

The assumption that the client is to some degree uninformed about the type of his own project is common in the auditing literature. Examples include the papers by Dye (1995), Dye (1995), Schwartz (1997), Chan / Pae (1998), Chan / Wong (2002) or Laux / Newman (2010). This assumption, although controversial, allows for misreporting by the client without assuming opportunistic behavior (e.g. earnings manipulation). Of course, the investing firm generally has some information about the prospects of the projects it intends to undertake. However, the assumption seems uncritical for several reasons: First, some information is implicitly captured by the assumption that the project net present value is positive, conditional on the additional information generated by the ICS and the audit. This implies that the firm has already pre-screened available investment opportunities. Second, it is not important that the ICS and the audit are the sole sources of information, but that the ICS and the audit add significant value through the additional information they provide. It might well be that the firm has some financial information about the project type, but this information can be sufficiently refined by the ICS and the external audit (this ex ante information would be reflected in the prior belief \(p\)). Third, it might even be that the firm is perfectly informed about the true project type, but is (to some extent) unable to translate this knowledge into accurate, GAAP-conform financial statement figures. Then, the ICS and the audit does not generate new information about the project, but checks whether the non-financial information correctly entered the firm’s financial reporting. Fourth, the project’s future prospects can steadily change. The project reporting might be overstated because the firm, despite being informed about the project type at the beginning, fails to keep up with unfavorable economic developments. Finally, if the prior probability \(p\) is greater than a half, then even in the absence of any additional information a good report is a legitimate best estimate. For all those reasons the firm, despite being uninformed along some dimension, does not act fraudulently when claiming that the project is of the good type.
auditing department, as well as the investment in technical equipment. For simplicity, I assume that the ICS quality $s$ equals the detection probability of a bad type project. With probability $(1 - s)$ a bad type project remains undetected by the internal controls, which corresponds to the control risk of the audit risk model. However, a project that is actually of the good type is never sorted out by the ICS, that is I exclude a type-I error.\footnote{An economic justification for the one-sided error assumption is that there exist some negative characteristics, by which bad type project can be clearly distinguished from good type projects. The existence of those characteristics is known by the client firm, but they can only be detected by an effective ICS. Hence, in my model the ICS is a technology that depending on the effort level $s$ can detect those characteristics more or less effectively. When the ICS processes a good type project, it can never react because the triggering bad-type-characteristics are not present. Examples for such characteristics are a non-trivial set of financial ratios, or a certain macroeconomic data constellation. Only a sufficiently effective ICS will be able to detect and report those circumstances. However, as will become clear in the comparative statics section, the introduction of a type-I ICS error would qualitatively not change any of the results.}

Installing and maintaining a high-quality ICS causes costs of $c(s) = bs^2$, whereby $b > 0$ is a constant cost parameter. After the installation, the ICS starts processing the project.
data and provides a result at the end of the period.

Next, an independent, risk-neutral auditor (she) is hired by the firm. Before actually auditing the financial report, the auditor assesses the firm’s ICS. In the engagement letter the auditor and the client agree upon fixing the audit fee $F$ after the auditor’s assessment of internal controls. This makes the fee contingent on the actual audit hours spent. The ICS assessment causes fixed costs of $W^{17}$ and perfectly informs the auditor about the state of the ICS prior to the actual audit procedures$^{18}$. After observing the state of the ICS, the auditor decides on the extent of the project audit procedures. Following Pae / Yoo (2001), I assume that the auditor can only observe the state of the ICS but not the actual outcome of the internal control procedures at the time she has to choose her audit effort. This modeling simplifies the analysis because it ensures that the auditor is always hired and has an incentive to exert positive audit effort. In appendix B I formally show that the results of the paper remain qualitatively unchanged if the client can observe the ICS outcome before he decides about hiring the auditor. Hence, I stick with the formally more simple base version with an unobservable ICS outcome in order to identify the interesting economic effects in better contour.

Subsequent to the assessment of the ICS, the auditor plans her substantive testing procedures. The audit market is competitive, that is the fee just covers her expected total costs. The auditor’s reservation utility is normalized to zero. The fact that the audit fee depends on the auditor’s ICS assessment is a salient difference to the model of Pae / Yoo (2001), in which the fee is fixed prior to the ICS assessment. This dependency allows the client to save on the audit fee by installing and documenting an effective ICS.

17 Internal control audits usually involve highly standardized procedures, such as checking the application of control models by checklists; see Abdolmohammadi / Wright (1987), p 5. Since I assume that the auditor audits the newly installed or adapted ICS for the first time, the extent of the ICS audit procedures and thus the ICS audit costs can be held constant.

18 According to Section 302 of the SOX, the management is legally obliged to report the state of the ICS to the auditor. Section 906 threatens heavy penalties if the manager misreports, which makes the perfect observability of $s$ seem plausible in the model. The perfect observability assumption is also made by Smith / Tiras / Vichitlekarn (2000) and Pae / Yoo (2001).
For the firm this practice is efficient because it enables potential cost savings. It is also what the PCAOB had in mind when it issued Auditing Standard No. 5 (AS5) in 2007 after massive complaints about the costs of complying with SOX Section 404. The main goal of AS5 was to bring down audit costs by allowing the auditors to rely on the work of others and thereby avoid testing redundancies. AS5 requires auditors to adopt a top-down approach, that is they shall adjust the amount of substantive testing procedures according to their prior assessment of internal control effectiveness. In the comments on the proposal the PCAOB and the SEC received, both auditors and firms highlighted the potential for cost-savings and efficiency gains.

After the fee is fixed, the auditor exerts unobservable audit effort \( a \in [0; 1] \) in order to issue an opinion on the veracity of the client’s high project report. The audit effort \( a \) also equals the detection probability of a wrongful report, respectively a bad type project. The probability \( 1 - a \) therefore corresponds to the detection risk factor in the audit risk model. Analogous to the ICS, the auditor never commits a type-I error.

Exerting audit effort causes direct costs of \( d(a) = ka^2 \), whereby \( k > 0 \) denotes a constant cost parameter.

For example, Procter & Gamble stated: 'We agree that the proposed standards have the potential to assist auditors in making their audit process more efficient and cost-effective'. Similar opinions were expressed by Microsoft, Vodafone PriceWaterhouseCoopers and KPMG. The collected responses to the PCAOB and the SEC can be accessed under https://pcaobus.org//Rulemaking/Pages/Docket021Comments.aspx. The effect of AS5 on audit fees has been empirically investigated by Krishnan / Krishnan / Song (2011) and Wang / Zhou (2012). Both papers find significant reductions in audit fees after the implementation of AS5.

In the base model, the client firm’s unaudited report is completely uninformative. In the setting with an observable ICS outcome, the high report is only submitted to the auditor, if the ICS did not detect a bad type. In this case, the high unaudited report is in fact based on acquired information about the project quality.

Ignoring a type-I error of the auditor is common in the auditing literature. Examples include Dye (1993), Schwartz (1997), Hillegeist (1999), Pae / Yoo (2001), Chan / Wong (2002), Liu / Wang (2006) or Laux / Newman (2010). Technically, allowing for a type-I error would not qualitatively influence the results, as long as the audit and control technologies are such that the aggregate probability of committing a type-I error is sufficiently small. If that were the case, the project would still be implemented only if the final audited report is high, because the probability of an understatement is considered to be rather insignificant. This argument is brought forward by Dye (1995), p 80. I also expect the results to remain qualitatively unchanged if (a) the audit and the control technology are such that the probability of both error types decreases in the players’ effort levels and (b) the liability consequences of an overstatement exceed the consequences of an understatement. Both conditions are considered to be realistic.
A low cost parameter $k$ represents a more sophisticated audit technology, or cheap, routine audit procedures.\textsuperscript{22}

At the end of the period the firm publishes a report about the project’s future prospects. This report is the joint product of the ICS and the audit and represents the firm’s audited financial statements. If at least one of the two monitoring institutions has detected a bad type project, the project is discarded. However, if neither the auditor nor the ICS have detected a bad type, the project is funded by the investors. Investors require a share $\alpha$ of the project’s returns for the provision of capital. The capital market is assumed to be competitive, that is $\alpha$ equals the share at which investors just expect to break even when providing $I$ after observing a favorable report. The reservation utility of all investors is normalized to zero. If ex post the project turns out to be of the good type, the project investors receive the share of returns $\alpha Z$, while the client keeps the remaining share $(1 - \alpha)Z$. However, with the probability $(1 - p)(1 - s)(1 - a)$ both the ICS and the auditor have failed to detect an overstatement and the investors lose $I$. Because there is a contractual financing relationship with the client, and a quasi-contractual relationship between the auditor and the project investors about the use of the financial report for making the financing decision, both the client firm and the auditor are jointly held liable for the loss of $I$. I assume the existence of a liability system that provides perfect investor protection. On the one hand, the liable parties always have the financial capacity to pay

\textsuperscript{22} In the model, internal controls and the audit are modeled as detection activities of the same kind. They only differ in terms of the cost parameters $b$ and $k$. I choose this simplifying assumption to work out the interaction between the ICS and the audit as interference-free as possible. In reality the tasks of the ICS and the auditor of course differ. The primary task of the ICS should be to increase the precision of financial statements, that is to avoid both over- and understatements (see also Ewert / Wagenhofer (2016a)). The auditor on the other hand, should be rather concerned about undetected overstatements, because they usually involve more severe investor losses than understatements. However, it can be shown that technological differences in the sense, that the ICS can commit both a type-I and a type-II error, whereas the auditor can only commit a type-II error do not change any of the comparative static results, as long as the type-I error consequences do not depend on the damage apportionment rule. In that case the term with the type-I error consequences vanishes when differentiating the first-order condition with respect to the damage apportionment parameter.
their share of damages in the model\textsuperscript{23}. On the other hand, the lawsuit is assumed to be costless for the investors. Assuming perfect protection of project investors simplifies the analysis because it suppresses a cost of capital-effect\textsuperscript{24}.

Given that the audited financial report claims a good project, some investors on the market invest an amount $H$ into firms that are economically related to the client firm because they are in the same industry or even in the same value-added chain. Both the scale and the success of investments $H$ are tied to the scale and the success of the investment $I$ into the client firm’s project. These additional investments are not as ambitious as the project of the client firm. With probability $p$ they yield returns with a present value of $H$, that is the net present value (NPV) of successful investments is normalized to zero. With probability $(1 - p)$ the investment $H$ is of the bad type and lost if undertaken.

Damages $I$ are assigned to the auditor and the client firm by an impartial court. The client firm has to pay damages in the amount of $tI$. The percentage factor $t \in [0; 1]$ hereafter is referred to as damage apportionment factor. I interpret $t$ as an expectation about how the court assigns damage payments. Although to my knowledge there are no countries with codified damage proportions, a legal tendency can be inferred from law and basic legal principles. Moreover, the auditor and the client firm might have formed an expectation about how the courts decide based on preceding cases\textsuperscript{25}. The remaining

\textsuperscript{23} The assumption that the firm needs external project financing but nevertheless has the capacity to pay its share of damages at the end of the period seems to apply for most investment situations. In reality, the decision about internal or external financing is mostly determined by differing capital cost rates. However, in the model I do not consider different costs of capital, which raises the question why the client seeks external financing at all. Conversely, one could ask, where the financial resources to compensate aggrieved investors come from, if the client by assumption does not have enough liquid assets to finance the project initially. I assume that the firm temporarily lacks of internal financing capacity because projects that have been started in the past yield returns only after the investment decision.

\textsuperscript{24} The assumption that total damages equal the investors’ losses (‘out-of-pocket losses’) simplifies the analysis and is common in the literature. See for example Narayanan (1994), Hillegeist (1999) and Lu / Sapra (2009). In appendix B I formally show that the results qualitatively go through if the total damages are assumed to be independent and smaller than the investors’ incurred losses.

\textsuperscript{25} Note that in reality the actual damage apportionment is likely to be endogenously influenced by an ex post court assessment of the respective effort levels. However, there will always be an underlying legal trend.
proportion of the investors’ damages \((1 - t)I\) has to be paid by the auditor.

Now consider the investors who lost \(H\) from their investments in related firms. Those investors cannot recover their loss from suing the client firm because there is no direct contractual financing relationship. Neither can they sue the firms they invested in, since those did not prepare the erroneous report. However, the auditor is aware of the economic connection between the client’s project and the other investments \(H\), and knows that her opinion is decisive for the investments \(H\). Therefore the investors of \(H\) are deemed foreseeable third parties and can sue the auditor for the breach of her public responsibility to protect them from incorrect financial information. By doing so they can expect to recover an amount \((1 - t)(r - 1)I\) of their investment. Obviously, the courts decision to allow those damage claims depends on two factors. First, it depends on the baseline legal tendency to hold auditors liable for investor losses. Second, it depends on the economic relation between the investments \(I\) and \(H\), that is the relevance of the audited report for the other investments. This is captured by the factor \((r - 1) < 1\). The greater the parameter \(r\) is, the tighter the economic bond between the two investments is. Because a successful investment yields a NPV of zero and the capital market is competitive, investors will not invest more than the amount they can recover in case of a loss. Therefore \(H = (1 - t)(r - 1)I\), which means that the unconditional expected NPV of the additional investments equals zero.\(^{26}\)

The model can be summarized by the expected utility functions of the players, denoted by \(U_C\) (client firm), \(U_A\) (auditor), \(U^I\) (investors of \(I\)) and \(U^H\) (investors of \(H\)):

\(^{26}\) Constraining the expected NPV of all other investments in the industry to zero keeps the model simple. Otherwise it would be necessary to impose additional assumptions about the investment possibilities within the industry. In section six I discuss why I consider this simplification to be made without loss of generality. Moreover, I restrict \(r\) to the interval \([1; 2]\) because I want the auditor’s liability concerns to primarily stem from the investment \(I\) of the firm she actually audits. This is reasonable to assume given that (i) the investment \(I\) is a pioneering, ‘bellwether’ project, and (ii) the liability towards the investors of \(I\) has a considerably stronger legal foundation than the liability for the loss of \(H\).
\[ U^C = p(1 - \alpha)Z - (1 - p)(1 - s)(1 - a)tI - \frac{hs^2}{2} - F, \quad (2.1) \]

\[ U^A = F - \frac{ka^2}{2} - (1 - p)(1 - s)(1 - a)(1 - t)rI - W, \quad (2.2) \]

\[ U^I = Pr(G | No)(\alpha Z - I) + (1 - Pr(G | No)) (I - I) = 0, \quad (2.3) \]

\[ U^H = Pr(G | No)(H - H) + (1 - Pr(G | No)) (1 - t)(r - 1)I - H = 0, \quad (2.4) \]

whereas \( Pr(G | No) = \frac{p}{p + (1 - p)(1 - s)(1 - a)} \) is the investors’ revised belief that the project is of the good type, if neither the ICS nor the auditor have detected a bad type. To ensure concavity of the client’s utility function \( \frac{d^2 U^C}{d a^2} < 0 \) and hence the existence of an interior solution, I assume that the control costs are sufficiently high, whereas \( b \) must exceed the threshold value

\[ \bar{b} = \frac{(1 - p)^2 r (1 - t) I^2 (r (1 - t) + 2t)}{k}. \]

2.4 The Equilibrium

The subgame-perfect equilibrium of the game is established by backward induction. If the investors decide to finance the project after having received a positive audited report, they set the cost of capital \( \alpha \) according to condition (2.3). After having observed the ICS state, the auditor chooses the equilibrium audit effort \( a^{*} \) that minimizes her total costs.

\[ \text{Assuming a lower bound for the control costs seems reasonable given the intense debate about the introduction of Section 404 of the SOX. Numerous SEC-listed companies and lobbies complained about the high costs caused by the additional requirements regarding the ICS.} \]

\[ \text{Hereafter, asterisks mark equilibrium variables.} \]
In a rational expectations equilibrium the client firm correctly anticipates $a^*$ and $\alpha$, and chooses the effort level $s^*$ that maximizes $U^C(s)$.

The equilibrium cost of capital rate can be directly obtained from condition (2.3) and equals $\alpha = \frac{I}{Z}$. The equilibrium audit effort is obtained by solving for the auditor’s first-order condition (FOC) $\frac{dU^A(a)}{da} = 0$ and reads

$$a^* = \frac{(1 - p)(1 - s)(1 - t)I}{k}$$ (2.5)

Lemma 2.1 gives two important properties of the equilibrium audit effort.

**Lemma 2.1.** The direct effects of a higher damage share $t$, and a higher ICS quality $s$ on the equilibrium audit quality $a^*$ are both negative.  

The economic intuition behind Lemma 2.1 is straightforward. If the auditor observes that the ICS is of high quality, she will cut down her effort because the marginal gain in assurance is low. In other words, auditing and internal controls are *strategic substitutes* in equilibrium. This substitution is the core element of the conventional multiplicative audit risk model. The auditor will also care less about liability and exert less effort if her damage share is smaller.

Before the auditor chooses her actual audit effort $a$, the client chooses the ICS quality $s$, knowing that the auditor can observe $s$. The equilibrium ICS quality $s^*$ is the level of $s$ that satisfies

$$\frac{dU^C}{ds} = (1 - a^*)(1 - p)tI + (1 - s)(1 - p)tI \frac{da^*}{ds} - \frac{dF}{ds} - bs = 0.$$ (2.6)

---

29 Derivation: See appendix A.
30 Proof: See appendix A.
31 Strategic substitution means that in a world with two input and one output good, increasing the amount of input good A decreases the *marginal* effect of input good B on the output; see Bulow / Geanakoplos / Klemperer (1985), pp 489 et seq.
**Remark 2.1.** For any feasible parameter tuple $\bar{v}$ from the parameter space $\mathcal{V} = \{p, k, b, t, r, I\}$, given the equilibrium audit effort $a^*$, there exists a unique ICS quality $s^*$ that maximizes the client’s expected utility $U^C$. \[32\]

It can be seen from equation (2.6) that the equilibrium ICS quality is determined by the net of several countervailing effects. On the one hand, a more effective ICS increases the probability of detecting a bad type project and thus raises the client’s expected utility. This direct effect is depicted by the term $(1 - a^*)(1 - p)tI > 0$ in equation (2.6). On the other hand, a higher ICS quality causes higher control costs, which are represented by the expression $-bs < 0$. However, the resulting net direct effect is only one side of the medal. Since the auditor makes the choice of the audit effort level dependent on the observed ICS state, the client has to take into account the influence of his ICS quality choice on the equilibrium audit effort. As stated in Lemma 2.1, a higher ICS quality incites the auditor to choose a lower audit effort in equilibrium ($\frac{da^*}{ds} < 0$). That indirectly raises the client’s expected liability costs, and thus negatively affects his expected utility. This negative partial effect is depicted by the product term $(1 - s)(1 - p)tI \frac{da^*}{ds} < 0$. These three effects are the ones that make up the model of Pae / Yoo (2001). In addition, the observability of the ICS quality prior to fixing the audit fee introduces a new fee effect, captured by the derivative

$$\frac{dF}{ds} = \frac{\partial F}{\partial s} + \frac{\partial F}{\partial a^*} \frac{da^*}{ds}. \quad (2.7)$$

Since the term $\frac{\partial F}{\partial a^*}$ is the auditor’s FOC and therefore must equal zero in equilibrium, the sign $\frac{dF}{ds}$ solely depends on the sign of the direct effect $\frac{\partial F}{\partial s}$, which is negative due to

$$\frac{\partial F}{\partial s} = -(1 - a^*)(1 - p)(1 - t)rI < 0.$$ 

32 Proof: See appendix A.
Hence, a higher ICS quality unambiguously lowers the audit fee. The reason for the
reduction is that a more effective ICS lowers the auditor’s expected liability payments. A
lower audit fee in turn positively influences the equilibrium ICS quality, as can be seen
from equation (2.6). The full form of the client’s FOC can now be rewritten as

\[
\frac{dU_C}{ds} = (1 - a^*)(1 - p)tI - bs + (1 - s)(1 - p)tI \frac{da^*}{ds} + (1 - a^*)(1 - p)(1 - t)rI = 0.
\]

(2.8)

2.5 Comparative Statics

2.5.1 ICS Quality

Suppose that in equilibrium \( t \) increases at the margin. This change is equivalent to an
increase of the client firm’s share of damages, respectively a relaxation of the auditor’s
liability\[^{33}\]. The most obvious legal measure triggering an increase in \( t \) is a court’s judgment
that assigns a surprisingly high (small) amount of the damages to the client firm (the
auditor). Following such a judgment, both auditors and firms will adapt their expectations
about the apportionment of damages in similar future cases. In a wider sense, all legal
measures that relatively alter the auditor’s and the client firm’s responsibility for accurate
reporting trigger a change in the damage apportionment factor \( t \). An increase in \( t \) also
captures the recent turning away from tight auditor liability in the U.S.

In equilibrium,

\[
\frac{d^2U_C}{ds dt} = 0 \iff \frac{ds^*}{dt} = -\frac{\frac{d^2U_C}{ds dt} + \frac{d^2U_C}{ds \partial a^*} \frac{\partial a^*}{\partial t}}{-\frac{d^2U_C}{ds^2}}.
\]

(2.9)

\[^{33}\] For the opposite measure, which is tightening of the auditor’s liability, respectively a relaxation of the
client firm’s liability, all of the following results exactly reverse.
Due to the strict concavity of $U^C$, the sign of $d\alpha^*/dt$ follows the sign of the numerator in (2.9). The full form of the numerator reads

\begin{equation}
K(t) = -[(1-p)tI + (1-p)(1-t)rI] \frac{\partial \alpha^*}{\partial t} + (1-s^*)(1-p)I \frac{\partial \alpha^*}{\partial s} \\
+ (1-p)(1-s^*)tI \frac{\partial^2 \alpha^*}{\partial s \partial t} + (1-a^*)(1-p)I(1-r).
\end{equation}

Analyzing this expression yields the following result:

**Proposition 2.1.** For a damage apportionment, which is sufficiently

(i) strict for the client firm ($t > \hat{t}_c$), any further increase of the client’s damage proportion $t$ strictly increases the equilibrium ICS quality $s^*$.

(ii) mild for the client firm ($t < \hat{t}_c$), any increase of the client’s damage proportion $t$ strictly decreases the equilibrium ICS quality $s^*$.$^{35}$

Part (i) of the proposition is in line with the reasonable conjecture that a harsher client firm liability increases the firm’s efforts to install a high quality ICS. However, part (ii) is unexpected. To understand the economic intuition, consider the expression $K(t)$ in (2.10). Obviously, the equilibrium reaction of the client firm to the liability shift $d\alpha^*/dt$ is driven by four different incentives. First, the liability shift directly reduces the auditor’s effort, which increases the likelihood of an audit failure. The client firm has an incentive to compensate the reduction of audit effort by increasing the ICS quality (effect (I)). Second, the client firm has a strategic incentive (effect (II)). Since it is the first-mover of the game and the auditor can observe the ICS quality before choosing her effort, the firm can influence $a^*$ by the choice of $s$. Due to the substitution between the two

$^{34}$ See the proof of Proposition 2.1 in appendix A for the derivation.
$^{35}$ Proof: See appendix A.
monitoring activities, the strategic incentive is negative, that is the firm has an incentive to decrease the ICS quality in order to incite the auditor to audit more carefully. Third, the firm takes into account that the power of this strategic leverage decreases as \( t \) increases, since the auditor basically cares less about liability if her damage share is small (effect (III)). Note that these are also the three incentives that drive the main result of Pae/Yoo (2001). The net effect of (I), (II) and (III) can be easily shown to be positive, that is a higher client damage share increases the equilibrium ICS quality. However, the negative effect (IV) is new. This effect represents a fee-decrease resulting from the liability shift, because a higher client damage share \( t \) reduces the expected liability to foreseeable third parties the auditor has to be compensated for via the fee.\(^{36}\) First, this fee-decrease lowers the equilibrium ICS quality, since the liability consequences lose weight in the client’s decision making. Second, the fee-effect perturbs the balance between the other three effects. For currently low values of \( t \), that is a sufficiently harsh auditor liability, the leverage effect (II) now becomes relatively powerful. Together with the fee-saving effect (IV) it then overwhelms the positive effects (I) and (III). Therefore in equilibrium the client firm intentionally reduces the ICS quality to incite the auditor to exert disproportionately more audit effort. However, for a sufficiently client-weighted damage apportionment \( (t > \hat{t}) \) the leverage loses power and the firm is better off by increasing the ICS quality. Figure 2.3 illustrates the client’s equilibrium choice of ICS quality for the parameter set \( b = 2, k = 3, p = 0.8, r = 1.3, I = 7. \)

### 2.5.2 Audit Quality

Proposition 2.2 summarizes how the auditor reacts to the liability shift \( dt > 0 \), which is equivalent to a relaxation of auditor liability.

\(^{36}\) It is obvious that without the additional auditor liability, effect (IV) would be zero, since the client does not care whether he pays the damage share \((1-t)/I\) to the auditor via the fee or directly to the investors.
Proposition 2.2. For a damage apportionment, which is sufficiently

(i) mild for the auditor \((t > \hat{t}_a)\), any further decrease of the auditor’s damage proportion \((1 - t)\) strictly decreases the equilibrium audit effort \(a^*\).

(ii) harsh for the auditor \((t < \hat{t}_a)\), any decrease of the auditor’s damage proportion \((1 - t)\) strictly increases the equilibrium audit effort \(a^*\).

(iii) The audit effort threshold value \(\hat{t}_a\) is strictly smaller than the ICS quality threshold value \(\hat{t}_c\).

Proposition 2.2 states that the relation between the auditor’s damage share and the equilibrium level of audit effort is ambiguous as well. To see what happens, consider the total derivative of \(a^*\) w.r.t. \(t\):

\[
\frac{da^*}{dt} = \frac{\partial a^*}{\partial t} + \frac{\partial a^*}{\partial s} \frac{ds^*}{dt}.
\]  

(2.11)

The direct effect \(\frac{\partial a^*}{\partial t}\) captures the auditor’s effort reduction due to her reduced liability share. The product in (2.11) captures the reaction to the client’s ICS adaption. For damage shares \(t\) slightly above the threshold value \(\hat{t}_a\), the direct effect \(\frac{\partial a^*}{\partial t}\) is rather

\[37\] Proof: See appendix A.
DAMAGE APPORTIONMENT AND ICS QUALITY

Figure 2.4: Equilibrium levels of audit effort $a^*$

strong, as the ICS quality already approaches its minimum and thus a higher amount of expected liability vanishes. At the same time, the strength of the strategic effect, which is depicted by the product $\frac{\partial a^*}{\partial s} \frac{ds^*}{dt}$, decreases as $t$ moves closer to $\hat{t}_c$ because the equilibrium ICS quality change $\frac{ds^*}{dt}$ becomes very small. Within the interval $t \in ]\hat{t}_a; \hat{t}_c]$ the rather strong direct effect outweighs the rather weak strategic effect of the client’s ICS adaption. For damage shares beyond $\hat{t}_c$, the strategic effect is also negative because the client increases the ICS quality in equilibrium. Hence, the equilibrium audit effort clearly decreases in $t$ for $t \in [\hat{t}_c; 1]$. Now consider part (ii) of Proposition 2.2. For a sufficiently large damage share of the auditor, the direct effect $\frac{\partial a^*}{\partial t}$ is rather weak, as the equilibrium ICS quality increases in the auditor’s damage share $(1-t)$ below $\hat{t}_c$. However, the leverage $\frac{\partial a^*}{\partial s}$ is rather strong because the auditor’s share $(1-t)$ is basically high. This means the auditor becomes less concerned about the direct impact of the damage apportionment on her liability consequences, but is motivated to exert more audit effort by the client firm’s strategic reduction of the ICS quality. Then relaxing a harsh system of auditors’ third party liability enhances audit quality. Figure 2.4 graphically illustrates the equilibrium audit effort, again for the parameter set $b = 2, k = 3, p = 0.8, r = 1.3, I = 7$.

38 It is easy to see that the strength of the direct effect strictly decreases in the equilibrium ICS quality due to $\frac{\partial^2 a^*}{\partial t \partial s} = \frac{(1-p)\epsilon I}{k} > 0$. 

48
2.5.3 Financial Reporting Accuracy

It now remains to put the pieces together and analyze the aggregate impact of a change of the damage apportionment factor on the total accuracy of financial reports. Let the accuracy of financial reports be defined as a function of the damage apportionment factor in the form

\[ Q(t) = 1 - (1 - p)(1 - s^*)(1 - a^*). \]

Then the following statement can be made about the effect \( \frac{dQ(t)}{dt} \):

**Proposition 2.3.** There exists an interval \( [t^l_Q; t^h_Q] \), whereas \( t^l_Q < \hat{t}_a \) and \( t^h_Q > \hat{t}_c \), in which the accuracy of audited financial reports \( Q(t) \) absolutely decreases in \( t \).\(^{39}\)

Proposition 2.3 states that extending the client firm’s proportion of damages \( t \) is detrimental for the accuracy of audited financial reports if the current liability distribution is sufficiently even. To see why, consider the total derivative of the financial reporting accuracy w.r.t. the damage apportionment factor:

\[ \frac{dQ(t)}{dt} = (1 - p) \left[ (1 - s^*) \frac{ds^*}{dt} + (1 - a^*) \frac{da^*}{dt} \right]. \]

At the threshold value \( \hat{t}_a \) the client reduces the ICS quality to react to the auditor’s direct effort reduction and the increase in his own liability share. However, the auditor does not do anything because the client’s effort stimulating reduction of \( s^* \) just outweighs the negative direct incentive of her liability reduction. It follows that a sufficiently small interval below the threshold \( \hat{t}_a \) exists, in which the auditor’s effort increase is too small to overcome the negative effect of the ICS quality reduction on the financial reporting

\(^{39}\) Proof: See appendix A.
DAMAGE APPORTIONMENT AND ICS QUALITY

accuracy. At the threshold value \( \hat{t}_c \) the client firm is indifferent between exploiting the leverage and increasing the financial reporting accuracy on its own. Thus the strategic effect on the audit effort is zero. What remains is the negative direct effect of the liability shift on the audit. Consequently a sufficiently small interval exists where the client’s equilibrium effort increase is too weak to overcome the negative effect of the equilibrium audit effort reduction on the accuracy of published reports. It is straightforward that between the threshold values, where both effort levels strictly decrease in \( t \), the overall accuracy of published reports must also decrease.\(^{40}\)

Now is a good point to oppose the results derived so far to the existent empirical evidence. For sufficiently high damage apportionment factors \( t > \hat{t}_c \), a more (less) client-weighted damage apportionment leads to an increase (decrease) in internal control quality and a decrease (increase) in audit quality. This equilibrium substitution is intuitive and is clearly supported by prior theoretical work (Nelson / Ronen / White (1988), Smith / Tiras / Vichitlekarn (2000) or Pae / Yoo (2001)). However, the overall empirical evidence on the relation between firm-side accounting activities and external auditing is very mixed. Numerous papers back the substitution hypothesis (Felix / Gramling / Maletta (2001), Bedard / Johnstone (2004), Raghunandan / Rama (2006), and Hogan / Wilkins (2008)), relying on the supply-side argument that better corporate governance reduces the marginal gain in assurance through the audit, and thus makes the auditor cut down her effort. However, more recent studies point out that the observable equilibrium relation between firm-side accounting and external auditing is rather complementary (Goodwin-Stewart / Kent (2006), Knechel / Willekens (2006), Hay / Knechel / Ling (2008)). This is exactly what happens within the intermediate range of damage apportionment factors \( t \in ]t_Q^L; t_Q^H[ \). The common denominator of the ”complementarity-view” papers is that they build on a demand-side argument, that is investing in internal controls

\(^{40}\) Note that it is not possible to make general statements about the development of the financial reporting accuracy outside the interval \( ]t_Q^L; t_Q^H[ \), as it is unclear which effect dominates in the outer domain ranges. Note furthermore, that for \( r > 2 \) this result reverses, as the Propositions 2.1 and 2.2 reverse.
increases the demand for external auditing by the firm and its stakeholders. For example, Goodwin-Stewart / Kent (2006) argue that firms commit to strong corporate governance by investing in a powerful package of internal and external auditing. Knechel / Willekens (2006) and Hay / Knechel / Ling (2008) propose that investments in internal controls shift the costs and benefits of better controls between different groups of stakeholders, which lead to an increased demand for external auditing. However, I argue that it is the presence of differential liability consequences, respectively the importance of the audited report for the whole industry, that shifts the firm’s demand for audit quality and overall financial reporting accuracy. The firm then invests into the internal control system, such that the auditor’s supply of audit quality equals the firm’s demand. In contrary to the arguments of the empirical studies listed above, this argument simultaneously considers the supply- and the demand-side. A testable implication of Propositions 2.1 to 2.3 is that in less regulated environments where duties are more vaguely defined and it is thus equally easy to sue the auditor and the client firm, we should observe a complementary relation, whereas in countries with sharply regulated internal control duties the relation should be one of substitution. In fact Knechel / Willekens (2006) document a complementary relation for Belgian firms whose corporate governance is weakly regulated. However, when internal controls are mandatory and control costs are high, they find that external auditing and internal controls become substitutes. Similarly, Hay / Knechel / Ling (2008) document a complementary association using data from New Zealand, which they consider to be a relatively unregulated environment.

What is also interesting about Proposition 2.3 is that the equilibrium complementarity evolves in a setting in which firm-side accounting and auditing are always strategic substitutes ($\frac{\partial s}{\partial a} < 0 \forall t \in [0; 1]$). Despite this is an interesting issue per se, I do not elaborate on it further in this paper. The effects of strategic substitution and complementarity are investigated in detail in the third dissertation paper.
2.5.4 Social Welfare

In this subsection I analyze whether an efficient equilibrium is implementable, and how the equilibrium social welfare varies with the damage apportionment. Social welfare is defined as the equally weighted sum of all players’ individual utilities. The only player who receives a rent in equilibrium is the client firm. Therefore social welfare equals the expected utility of the client as given in equation (2.1), and fully stated reads

\[ U^{SW} = p(1-\alpha)Z - (1-p)(1-a)(1-s)tI - \frac{bs^2}{2} - (1-p)(1-s)(1-a)(1-t)rI - \frac{ka^2}{2} - W. \]  

As in the base model, I assume a lower bound for the control costs \( b \) which ensures that the social welfare function is strictly concave in its arguments \( a \) and \( s \). If the regulator could effectively enforce desired levels of audit and ICS quality, he would choose \( s \) and \( a \) such that (2.12) is maximized. Solving the FOC \( \nabla U^{SW}(a,s) = 0 \) for \( a \) and \( s \) gives the efficient levels of audit- and ICS quality, which are per se not interesting. However, what is interesting is the following finding:

**Proposition 2.4.** The equilibrium allocation of ICS- and audit quality is Pareto-efficient if and only if the auditor bears all the liability \((t = 0)\).

The intuition behind this result is as follows: For any damage apportionment \( t > 0 \), the auditor does only care about the fraction \((1 - t)I + H\) of the investors’ damages when choosing the audit effort. The client firm however always cares about the total amount of damages \((I + H)\) because it has to fully compensate the auditor for the damage share \((1 - t)\) via the fee anyways. Therefore, from the client’s perspective the audit effort is

---

42 It can be shown that the threshold value \( b \) of the base model is not high enough to make both the numerator and the denominator become positive in all cases.

43 Proof: See appendix A.
always too low for a non-zero \( t \), given that the equilibrium \( s^* \) equals the efficient ICS quality. This is why he has an incentive to help a bit by exploiting the strategic effect \( \frac{ds^*}{ds} \). This behavior distorts the equilibrium. However, for \( t = 0 \), the auditor’s and the client firm’s liability incentives are fully aligned, and thus there is no incentive for the client to strategically adjust the audit effort via the choice of ICS quality. Both players consider the same amount of (auditor) liability in their decision (the auditor considers it as direct liability costs, and the client considers it via the fee), and balance these costs against the direct effort costs. This is exactly what the social planner would do when maximizing social welfare. Because the strategic interaction is muted, the sequence of events does not matter. Therefore the simultaneous choice of the social planner yields the same solution as the equilibrium that evolves out of the dynamic game.

This result is intriguing in the light of prior theoretical research. First, it contrasts the finding of Pae / Yoo (2001), who claim that the efficient allocation of audit- and ICS quality cannot be implemented by choosing the amount of auditors’ liability payments.

The reason why an efficient allocation is indeed achievable in my model is the fact that the fee set after the auditor’s ICS assessment and therefore dependent on the client’s choice of ICS quality, which is why there cannot be an alignment via the fee. Second and more important, Proposition 2.4 provides a justification for the common ‘deep pocket’ assumption. The majority of papers on auditors’ liability simply assumes that the auditor bears all the liability, whereas the client firm remains without sanctions. However, given that audit fees are dependent on the actual audit effort, this extreme damage apportionment leads to a Pareto-efficient equilibrium. Hence, the deep-pocket assumption is not only a mathematical simplification, but has a solid

\[44\] It shall be underlined that efficiency is only achieved with regard to the equilibrium allocation of audit and ICS quality. This means that for \( t = 0 \), the equilibrium set by a social planner and the equilibrium of the game coincide because the moral-hazard problem effectively disappears. However, as I will show in a minute, \( t = 0 \) is generally not the welfare-maximizing damage apportionment, that is the second-best welfare for \( t > 0 \) is likely to exceed the first-best welfare in the \( t = 0 \)-game.
econmic foundation.

While Proposition 2.4 is interesting with regard to prior research, a regulator should rather be interested in choosing a damage apportionment rule such that the equilibrium welfare is maximized, even if this means that the moral-hazard problem of inducing the auditor to exert effort renders the equilibrium inefficient. Differentiating the equilibrium social welfare $U^{SW}(a^*,s^*)$ w.r.t. $t$ generally gives

\[
\frac{dU^{SW}(a^*,s^*)}{dt} = \frac{\partial U^{SW}(a^*,s^*)}{\partial t} + \frac{dU^{SW}(a^*,s^*)}{ds} \frac{ds}{dt} + \frac{\partial U^{SW}(a^*,s^*)}{\partial a} \frac{\partial a^*}{\partial t} \tag{2.13}
\]

The second addend is zero because its first factor equals the client firm’s FOC. Analyzing the remaining parts of the derivative (2.13) leads to the following proposition:

**Proposition 2.5.** (i) For sufficiently small direct audit costs $k < \hat{k}$, there exists a single positive damage apportionment factor $t^* \in [0; 1]$ such that the equilibrium social welfare $U^{SW}(a^*,s^*)$ is maximized.

(ii) For sufficiently high direct audit costs $k > \hat{k}$, the equilibrium social welfare $U^{SW}(a^*,s^*)$ is maximized if the client firm bears all the liability ($t^* = 1$).

Equation 2.13 shows that two, yet well-known effects determine the impact of damage apportionment on social welfare. On the one hand, a liability shift towards the client firm enhances welfare because it reduces the expected extra-liability towards foreseeable third parties the auditor has to be compensated for via the fee. This positive effect is captured by the partial derivative $\frac{\partial U^{SW}(a^*,s^*)}{\partial t}$. On the other hand, the liability shift $dt > 0$ decreases the auditor’s liability consequences and thereby weakens her incentive to exert high effort. This negatively affects welfare because it increases the chance that bad type
projects get financed. This negative impact is grasped by the product \( \frac{\partial U^{SW}(a^*, s^*)}{\partial t} \) in equation (2.13). If the audit costs are below the threshold value \( \hat{k} \), then the negative effort effect outweighs the positive savings effect for sufficiently high \( t > t^* \). The reason for this relation is that for low direct audit costs \( k \) the equilibrium audit effort \( a^* \) is considerably high and also highly reactive to changes in \( t \). Hence, the probability of litigation and thus the marginal savings in extra-liability are rather small compared to the detrimental marginal effect of an audit effort reduction. It follows that for \( k < \hat{k} \), a regulator should set \( t = t^* > 0 \) to maximize social welfare. However, if the direct audit costs are sufficiently high, the audit effort is basically lower and less reactive. Then the reduction of the deadweight loss outweighs the effort reduction effect for all possible damage apportionment rules. In this case social welfare is maximized, if the client bears all the liability. Figure 2.5 shows the equilibrium social welfare for the parameter sets \( b = 3, p = 0.8, r = 1.3, I = 7, Z = 10, W = 1, k = 2.5 \) (case (i)) and \( k = 4 \) (case (ii)) in the first-best case \( U^{SW}(a^f, s^f) \) and in the second-best case \( U^{SW}(a^*, s^*) \).

Proposition 2.5 holds ready the following policy recommendations: The higher the direct audit costs within an economy are, the more of the liability should be loaded onto the client firm. Conversely, if the average audit costs in an economy are low, then a harsher
auditor liability is necessary to implement the welfare-optimal equilibrium. This result is interesting in the light of the latest technological advances in data-processing - buzzwords 'big-data' and 'full audit'. The benefits for audit firms associated with cost-reducing innovations might enable a more even damage apportionment without encountering adverse welfare effects. Conversely, as the audit profession faces new challenges, such as blockchain, crypto-currencies and the use of artificial intelligence, more responsibility should be shifted to the reporting firms.

2.6 Robustness of the Results

I adopt several simplifying assumptions to keep the analysis tractable. The driving force of the model’s main results is the fact that the auditor’s extra-liability towards foreseeable third parties depends on her damage share. I consider this assumption to be realistic, because in legal environments where the courts consider the auditor to be more guilty than the client, capital market participants who made unfavorable decisions based on a wrong audit opinion are more tempted to jump on the bandwagon of auditor liability. Furthermore, the public opinion about the auditor’s breach of duty might be higher if she is assigned a higher proportion of the damages. However, one might argue that the propensity to sue the auditor depends more on the total size of the accounting scandal (the amount of bad investment the auditor has failed to prevent) than on the proportional damages. Adding such ”share-independent” costs \(-rI\) to the auditor’s expected utility function would not qualitatively change the model results, as long as a part of the extra-liability depends on the damage share. Indeed the possibility of share-independent extra-liability concerns makes the model assumption \(r < 2\) less restrictive.\(^{47}\) Note furthermore, that the results do not depend on the specific (linear) relation. It is only important that

\(^{47}\) Looking at the news on recent accounting scandals, it shows that in some cases the size of the accounting scandal and in some cases the auditor’s liability payment or penalty is reported. Which type of disclosure drives investors’ suing behavior more, is an empirical question.
some extra-losses of the auditor are monotonously increasing in the damage share \((1-t)\).

Another noteworthy point is that \(r\) is an economy-specific factor, not an auditor-specific one. If *ceteris paribus* the firm had the option to choose among auditors with different \(r\)’s it would choose the one with the lowest \(r\), since \((1-t)(r-1)I\) is a deadweight loss it has to pay.

Another discussion point is the assumption that the other investments even in case of success yield a zero NPV. Instead, one could assume that a sufficiently small proportion of the other investments \(H\) yields a positive NPV \(Y > I\) in case of success, whereas the rest yields zero NPV. In expectations the NPV of a successful investment would then be positive. Maintaining the assumption of a competitive capital market, the additional investment \(H\) would still be an increasing function of the auditors damage share \((1-t)I\), which is sufficient to obtain the main results. Moreover, because only a sufficiently small number of positive NPV projects is available, any variation in other firms’ investment level regards the zero NPV projects. Thus, the expected surplus generated by all other investment is invariant in the damage apportionment, and the results of the welfare analysis qualitatively remain unchanged as well.

At first glance it seems odd that the client does not receive the ICS signal before he hires the auditor. Compared to a setting where it is possible to delay the start of the audit until the ICS signal becomes available, this is detrimental because if the ICS could sort out a bad project, the audit would actually not be necessary. In such a modified setting the client basically has an additional incentive to increase the ICS quality, because he thereby increases the probability of avoiding an unnecessary audit. However, in equilibrium the auditor is still hired with a positive probability and this game-branch has to

---

48 This can be seen from equation (2.10). As long as the auditor’s extra-loss is decreasing in \(t\), the effect \((IV)\) will be negative.

49 Strictly speaking, this requires that the investors of \(H\) do not negotiate the cost of capital with the firms they invest in, but decide upon the investment level given their expectation about the future return. This implies that the financing relationship is less close than the one between the investors of \(I\) and the client firm. Alternatively, \(H\) might also be direct investments of other firms.
be considered by the firm ex ante. Therefore the basic "make-or-buy" rationale remains intact. In appendix B, I formally analyze this extension and show that the results do not qualitatively change if the client can make the auditor hiring decision contingent on the ICS signal.

Next, the model only pictures the strategic interaction between the auditor and the client firm. Because the investors receive full compensation in case of an audit failure, the lawsuit is risk-free and thus always initiated. Relaxing this assumption, for example by introducing a risky lawsuit, would require to additionally consider the investors’ suing decision. Presumably, this would implant additional interesting interdependencies into the model, but without doubt at the cost of further mathematical complexity. If the suing decision were non-strategic, such as in Hillegeist (1999) or Laux / Newman (2010), investors would receive a damage compensation lower than their initial investment provision \((D < I)\). In that scenario the results would not qualitatively change.\(^{50}\) The client firm would be additionally interested in keeping the cost of capital low by publishing accurate financial reports and thereby avoiding bad investment. Then the two negative addends in equation (2.10) would become accompanied by a third one, capturing the client firm’s incentive to lower the cost of capital by leveraging the audit effort through the ICS. This effect would enlarge the range of low damage apportionment factors for which the leverage effect is strong enough to make the conscious weakening of the ICS favorable. The case \(D < I\) also captures the case that the client can only partially compensate the investors for their losses, whereas the auditor’s pockets are deep.

Another model characteristic that calls for a discussion is the exogeneity of the damage apportionment factor. In reality, the actual damage apportionment is expected to depend on both the ICS quality and the auditor’s audit effort. To include an endogenous (= effort-dependent) damage apportionment, one would have to model the court’s decision

\(^{50}\) See appendix B for an analytical proof of this proposition.
by introducing a damage assignment function $t(s, a)$. This function should reasonably have the properties $\frac{\partial t(s,a)}{\partial s} < 0$ and $\frac{\partial t(s,a)}{\partial a} > 0$, that is the client’s damage share increases in the positive difference between the audit and the ICS quality. Narayanan (1994) first showed that such endogeneity of the actual liability payments creates additional, either positive or negative effort incentives, because the equilibrium effort choices are then not only determined by the actual damage proportion $t$, but also by the sensitivity of the damage proportion to the effort levels. Maintaining the interpretation of $t$ as a general jurisdictional tendency, respectively an expectation about the court’s judgment, one could assume the assignment function $t(a, s) = t + v(s, a)$, where $v$ is the exogenous legal trend and $v(a, s)$ is the endogenous component that has the aforementioned properties. Performing comparative statics for $t$ would qualitatively give the same results as the base model, as long as the superimposing sensitivity effects $\frac{\partial v(s,a)}{\partial (\cdot)}$ are not too strong. Overall, the extension would seed considerable additional mathematical complexity into the model without qualitatively changing the mechanism showed here. Moreover, Narayanan (1994) shows that it is possible to design a court assessment function, such that the liable parties exert the same effort as under the more rudimentary, exogenous regime. Hence, for any equilibrium it should be possible to design an apportionment mechanism, such that the second-order effects of the legal regime cancel out. By the same argument the simplifying assumption of a strict liability regime can be justified, despite in reality the negligence regime prevails.

51 The present model can be seen as a special case of this extension where $v(a, s) = 0$. The same argument is used by Hillegeist (1999) to justify the exogeneity of the damage apportionment rule.

52 Furthermore, it would be necessary to impose a series of limiting regularity conditions on the damage apportionment mechanism to ensure that the actual damage shares add up to one while having different slopes (and curvatures).

53 Narayanan (1994) formally proves this for the special case of a comparison between the joint-and-severalliability rule ($t = 1$ in case of client bankruptcy) and the proportionate liability rule ($t < 1$). This result is basically not surprising, since allowing for an additional degree of freedom cannot restrict the scope of action.

54 See Narayanan (1994) for the formal analysis of this issue.
2.7 Conclusion

This paper studies the interplay of damage apportionment, ICS and audit quality, financial reporting accuracy, and social welfare. The analysis takes place in a setting where auditors face potential additional liability towards capital market participants other than the primary addressees of the audit report due to their legal mandate in investor protection. In contrast to prior findings, I argue that the relation between the damage apportionment rule and the effectiveness of internal controls, the accuracy of financial reports and social welfare is non-monotone. Basically, the decision problem the client firm faces when the damage apportionment changes, is the choice between 'producing' more accurate financial reports on its own or 'outsourcing' this process to the auditor. Which strategy is favorable, critically depends on the strength of the leverage that the client firm can apply on the auditor. Because the equilibrium ICS quality steers the auditor's behavior, the equilibrium audit quality is also ambiguous in the damage apportionment. For currently harsh auditor liability, a reduction of the auditor's damage share may indeed increase the equilibrium audit quality. Putting these results together, it turns out that auditing and internal controls can be both substitutes and complements, as it is suggested by the mixed empirical evidence. The effect of damage apportionment on social welfare turns out to be ambiguous and dependent on the cost-efficiency of the audit technology available. Interestingly, it shows that a regulator can actually implement a Pareto-efficient equilibrium by assigning all damages to the auditor. This finding economically justifies the common assumption that the auditor pays all the damages out of her 'deep pockets', while the client firm remains without sanctions.

The main point I want to make with this paper is that the effective damage apportionment within a proportionate liability regime unfolds distinct economic forces as soon as multiple defendants are subject to different liability regimes. Because the auditor's responsibility transcends her triangular relationship with the client firm and its investors, holding the auditee liable for investor losses is not equivalent to holding the auditor liable. Holding
the auditor liable opens a valve in the auditor-client-investor triangle through which the auditors’ responsibility towards other industry firms and their investors spills back into the production-process of the auditee’s financial statements. By varying the damage share of the auditor, the regulator controls the flow through this valve. Although the analysis here is tailored towards financial statement audits, the results hold true for all situations in which two or more parties are responsible for a joint outcome, but face different liability regimes in case of failure. Examples include the sharing of liability among general and limited partners of a partnership, or the distinction between the product liability of a manufacturer and the statutory warranty of a retailer.
2.8 Appendix A

Derivation of the equilibrium audit effort

Differentiating the auditor’s utility function (2.2) w.r.t. the audit effort and setting the derivative equal to zero yields

$$\frac{dU^A}{da} = -ka^* + (1 - p)(1 - s)(1 - t)rI = 0.$$

Rearranging and solving for $a^*$ gives

$$a^* = \frac{(1 - p)(1 - s)(1 - t)rI}{k}.$$

Proof of Lemma 2.1

The partial derivatives of $a^*$ read as follows:

$$\frac{\partial a^*}{\partial t} = \frac{-(1 - p)(1 - s)rI}{k} < 0,$$

$$\frac{da^*}{ds} = \frac{-(1 - p)(1 - t)rI}{k} < 0.$$

QED.

Derivation of the equilibrium control effort $s^*$

Inserting $a^* = \frac{(1 - p)(1 - s^*)(1 - t)rI}{k}$ into the client’s FOC (2.8) and simplifying gives
\[
(1 - \frac{(1 - p)(1 - s^*)(1 - t) rI}{k})(1 - p) t I - bs^* - \frac{(1 - p)^2(1 - s^*) t(1 - t) r I^2}{k} + \frac{(1 - (1 - p)(1 - s^*)(1 - t) r I}{k}) (1 - p) (1 - t) r I = 0 \Leftrightarrow \\
(1 - \frac{(1 - p)(1 - s^*)(1 - t) r I}{k}) [(1 - p) I(t + r(1 - t))] - bs^* - \frac{(1 - p)^2(1 - s^*) t(1 - t) r I^2}{k} = 0 \Leftrightarrow \\
\Leftrightarrow \frac{1}{k} \left\{ [k - (1 - p)(1 - t) r I + s^* (1 - p)(1 - t) r I] [(1 - p) I(t + r(1 - t))] - b k s^* - (1 - p)^2 tr I^2 (1 - t) + s^* (1 - p)^2 tr I^2 (1 - t) \right\} = 0 \Leftrightarrow \\
\Leftrightarrow [k - (1 - p)(1 - t) r I] [(1 - p) I(t + r(1 - t))] - (1 - p)^2 tr I^2 (1 - t) \Leftrightarrow \\
b k s^* - s^* (1 - p)(1 - t) r I [(1 - p) I(t + r(1 - t))] - s^* (1 - p)^2 tr I^2 (1 - t) \Leftrightarrow \\
\Leftrightarrow s^* = \frac{[k - (1 - p)(1 - t) r I] [(1 - p) I(t + r(1 - t))] - (1 - p)^2 tr I^2 (1 - t)}{b k - (1 - p)(1 - t) r I [(1 - p) I(t + r(1 - t))] - (1 - p)^2 tr I^2 (1 - t)} \Leftrightarrow \\
\Leftrightarrow s^* = \frac{(1 - p) I [(k - (1 - p)(1 - t) r I)(t + r(1 - t)) - (1 - p) tr I(1 - t)]}{b k - (1 - p)^2 tr I^2 [(1 - t)(t + r(1 - t))] - (1 - p)r I [2t + (1 - t) r]} \Leftrightarrow \\
\Leftrightarrow s^* = \frac{(1 - p) I [k(t + r - rt) - (1 - p)(1 - t)r I(2t + (1 - t) r)]}{b k - (1 - p)^2 tr I^2 [(1 - t)(t + r)] - (1 - p)r I [2t + (1 - t) r]} \Leftrightarrow \\
\Leftrightarrow s^* = \frac{(1 - p) I [k(t + r - rt) - (1 - p)(1 - t)r I(2t + (1 - t) r)]}{b k - (1 - p)^2 tr I^2 [(1 - t)(t + r) + 2t]}.
\]

(2.14)

**Proof of Remark 2.1**

The first derivative of \( U^M \) w.r.t. \( s \) can be written as

\[
\frac{dU^M}{ds} = (1 - a^*)(1 - p) t I + (1 - s)(1 - p) t I \frac{da^*}{ds} - \frac{dF}{ds} - bs,
\]

whereby the fee effect \( \frac{dF}{ds} \) reads

\[
\frac{dF}{ds} = -(1 - a^*)(1 - p)(1 - t) r I + \left[ ka^* - (1 - s)(1 - p)(1 - t) r I \right] \frac{da^*}{ds} = \\
- (1 - a)(1 - p)(1 - t) r I < 0.
\]

63
It is obvious that the first derivative is not strictly monotonous. Differentiating $U^M$ w.r.t. $s$ once more gives

$$\frac{d^2U^M}{ds^2} = -2(1-p)\frac{dI}{ds} - \frac{d^2F}{ds^2} - b,$$

whereby

$$\frac{d^2F}{ds^2} = (1-p)(1-t)\frac{da^*}{ds}.$$

The second derivative of $U^M$ w.r.t. $s$ then can be expressed as

$$\frac{d^2U^M}{ds^2} = -2(1-p)\frac{dI}{ds} - (1-p)(1-t)\frac{dI}{ds} - b.$$

Inserting for $\frac{dI}{ds}$ and factoring out, finally yields

$$\frac{d^2U^M}{ds^2} = -(1-p)^2(1-t)\frac{dI^2}{ds} - \frac{rI^2}{k} - b.$$ 

Because the first term equals the assumed lower bound $b$ of the control costs, it follows that $\frac{d^2U^M}{ds^2} < 0$, which means that $s^*$ is the global maximum.

QED.

**Proof of Proposition 2.1**

Implicit differentiation of the client’s FOC w.r.t. the damage apportionment factor $t$ generally gives

$$\frac{d^2U^M}{dsdt} = \frac{d^2U^M}{dsdt} + \frac{d^2U^M}{ds\partial a} \frac{\partial a^*}{\partial t} + \frac{d^2U^M}{d^2s} \frac{d^2s}{dt} = 0.$$
This condition can be rearranged to

\[
\frac{ds^*}{dt} = \frac{\frac{d^2U}{dsdt}}{\frac{d^2U}{ds^2} + \frac{\partial^2a^*}{\partial s^2} dt}.
\]  

(2.15)

Because the derivative \( \frac{d^2U}{ds^2} \) is negative for all \( b > b_0 \), the denominator of (2.15) is positive. Hence, the sign of \( \frac{ds^*}{dt} \) is equal to the sign of the numerator of (2.15). The full form of the numerator reads

\[
K(t) = (1 - a^*)(1 - p)I + (1 - s^*)(1 - p)I \frac{\partial a^*}{\partial s} + (1 - s^*)(1 - p)tI \frac{\partial^2a^*}{\partial s\partial t} - (1 - a^*)(1 - p)(1 - t)rI \frac{\partial a^*}{\partial t} \Leftrightarrow
\]

\[
\Leftrightarrow K(t) = -[(1 - p)tI + (1 - p)(1 - t)rI] \frac{\partial a^*}{\partial t} + (1 - s^*)(1 - p)tI \frac{\partial^2a^*}{\partial s\partial t} + (1 - a^*)(1 - p)tI(1 - r) + (1 - s^*)(1 - p)tI \frac{\partial a^*}{\partial s} \Leftrightarrow
\]

\[
\Leftrightarrow K(t) = -[2(1 - p)tI + (1 - p)(1 - t)rI] \frac{\partial a^*}{\partial t} + (1 - a^*)(1 - p)I(1 - r) + (1 - s^*)(1 - p)I \frac{\partial a^*}{\partial s},
\]  

(2.16)

which, after inserting the partial derivatives \( \frac{\partial a^*}{\partial t} \) and \( \frac{\partial a^*}{\partial s} \), can be simplified to

\[
K(t) = \frac{-(1 - p)I[k(r - 1) + 2(1 - p)rI(1 - s^*)(-r(1 - t) - 2t + 1)]}{k}.
\]

It is obvious, that \( K(t) \) is negative (positive), and hence \( \frac{ds^*}{dt} \) is decreasing (increasing) if

\[
k(r - 1) > \frac{-2(1 - p)rI(1 - s^*)(-r(1 - t) - 2t + 1)}{-t > 0} \]

(2.17)

holds. Since the left hand side (LHS) of (2.17) is a constant, it is necessary to investigate
the right hand side (RHS), denoted by $V$. Differentiating $V$ w.r.t. $t$ gives

$$dV \frac{dt}{dt} = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial s} \frac{ds^*}{dt} \Leftrightarrow$$

At any intersection point between $V$ and the constant $k(r-1)$ the derivative $\frac{ds^*}{dt}$ is by definition zero. It follows that at any intersection point $V$ must have a positive slope in $t$, because $\frac{\partial V}{\partial t}$ is positive for $r < 2$. Due to the continuity of the functions, only one such intersection domain $\hat{t}_c$ can exist, since any further intersection domain would require $V$ to have a negative slope at this domain, and thus yield a contradiction. However, an intersection at a domain where $V$ is positively sloped does not yield a contradiction. To see why, consider the area around $\hat{t}_c$. Below $\hat{t}_c$ the product $\frac{\partial V}{\partial s} \frac{ds^*}{dt}$ is positive anyways because $\frac{ds^*}{dt}$ is negative. Slightly above $\hat{t}_c$, the product $\frac{\partial V}{\partial s} \frac{ds^*}{dt}$ is negative, but too weak to overcome the positive derivative $\frac{\partial V}{\partial t}$. Because $V$ can be shifted up- or downwards independently of the constant $k(r-1)$ by varying other parameters such as $b$, it does not asymptotically approach $k(r-1)$. Hence, the threshold value $\hat{t}_c$ always exists and it is only a matter of the parameter choice to ensure that it lies within the interval $[0; 1]$. To sum up, for all $t < \hat{t}_c$, $k(r-1) > V \Rightarrow K(t) < 0$ holds, and hence the derivative $\frac{ds^*}{dt}$ is negative. For all $t > \hat{t}_c$, $k(r-1) < V \Rightarrow K(t) > 0$ holds, and hence the derivative $\frac{ds^*}{dt}$ is positive.

QED.

**Proof of Proposition 2.2**

Consider the auditor’s equilibrium reaction $\frac{ds^*}{dt}$, which reads

55 It is obvious from the explicit expression in (2.14) that $s^*$ is strictly monotonously decreasing in $b.$
\[
\frac{da^*}{dt} = \frac{\partial a^*}{\partial t} + \frac{\partial a^*}{\partial s} \frac{ds^*}{dt} > 0
\]

From Proposition 2.1 it is known that \(\frac{ds^*}{dt}\) is strictly positive for \(t > \hat{t}_c\). Hence, \(\frac{da^*}{dt}\) is clearly negative for all \(t > \hat{t}_c\). It remains to investigate the behavior of \(\frac{da^*}{dt}\) in \(t\) for \(t < \hat{t}_c\).

Differentiating \(\frac{da^*}{dt}\) once more w.r.t. \(t\) gives

\[
\frac{d^2a^*}{dt^2} = \frac{\partial^2a^*}{\partial t^2} + \frac{\partial a^*}{\partial t} \frac{ds^*}{dt} + \frac{\partial a^*}{\partial s} \frac{d^2s^*}{dt^2}
\]

whereas

\[
\frac{\partial^2a^*}{\partial s \partial t} = \frac{\partial^2a^*}{\partial t \partial s} = \frac{(1 - p)I}{k} > 0,
\]

and \(\frac{\partial^2a^*}{\partial t^2} = 0\). Hence (2.18) simplifies to

\[
\frac{d^2a^*}{dt^2} = 2 \frac{\partial a^*}{\partial s} \frac{ds^*}{dt} + \frac{\partial a^*}{\partial s} \frac{d^2s^*}{dt^2}.
\]

Now let \(-\frac{dV_M}{ds^*} > 0\) (the denominator of (2.15)) be denoted by \(D(t)\). The derivative \(\frac{d^2s^*}{dt^2}\) can now be compactly written as

\[
\frac{d^2s^*}{dt^2} = \frac{dK(t)}{dt} D(t) - K(t) \frac{dD(t)}{dt} \frac{dK(t)}{dt} (D(t))^2.
\]

From the proof of Proposition 2.1 it is already known that \(K(t)\) is negative for all \(t < \hat{t}_c\). Furthermore, it follows from \(\frac{dV}{dt} > 0\) for \(t < \hat{t}_c\) that \(\frac{dK(t)}{dt}\) is strictly positive within that range. Because
\[
\frac{dD(t)}{dt} = \frac{(1 - p)^2 r I^2}{k} [r(1 - t) + 2t + r(1 - t) - 2(1 - t)] \Leftrightarrow \frac{(1 - p)^2 r I^2}{k} [2(1 - t)(r - 1) + 2t] > 0,
\]

the second derivative \(\frac{d^2 s^*}{dt^2}\) is positive for \(t < \hat{t}_c\). It follows that \(\frac{d^2 a^*}{dt^2}\) is clearly negative for \(t < \hat{t}_c\).

To prove that a finite threshold value \(\hat{t}_a\) exists, it remains to show that \(\frac{da^*}{dt}\) does not asymptotically approach zero, but can become positive for sufficiently small values of \(t\). Note that \(\frac{d^2 s^*}{dt^2}\) for \(t < \hat{t}_c\) means that the negative slope \(\frac{ds^*}{dt}\) is steeper for lower values of \(t\). Furthermore, it is already known that \(\frac{\partial a^*}{\partial s}\) is more negative for smaller values of \(t\).

Hence, the positive product reaches a local maximum at \(t = 0\). It is also known that the negative effect \(\frac{\partial a^*}{\partial t}\) does not directly depend on \(t\), but strictly increases and therefore becomes weaker as \(s^*\) increases. Since \(s\) strictly increases for \(t < \hat{t}_c\) as \(t\) becomes smaller, the effect \(\frac{\partial a^*}{\partial t}\) reaches a local maximum at \(t = 0\). This means that its absolute value has a local minimum at \(t = 0\). It can be concluded that for certain parameter constellations an interval \([0; \hat{t}_a]\) exists in which the positive product of the rather powerful effects \(\frac{\partial a^*}{\partial s}\) and \(\frac{ds^*}{dt}\) outweighs the rather weak negative effect \(\frac{\partial a^*}{\partial t}\) \(\overset{56}{\text{[56]}}\) It is straightforward that the threshold value \(\hat{t}_a\) where

\[
\frac{\partial a^*}{\partial t} = -\frac{\partial a^*}{\partial s}\frac{ds^*}{dt}
\]

holds must be smaller than \(\hat{t}_c\), since \(\frac{ds^*}{dt}\) must be negative for the condition to be fulfilled.

QED.

\(\overset{56}{\text{[56]}}\) It is easy to find such parameter constellations that also fulfill the other model requirements.
Proof of Proposition 2.3

The derivative of $Q(t)$ w.r.t. $t$ reads

$$\frac{dQ(t)}{dt} = (1 - p) \left[ (1 - a^*) \frac{ds^*}{dt} + (1 - s^*) \frac{da^*}{dt} \right].$$

Within the interval $] \hat{t}_a; \hat{t}_c[$ both $\frac{da^*}{dt}$ and $\frac{ds^*}{dt}$ are negative. Hence, $\frac{dQ(t)}{dt}$ is clearly negative.

At the threshold value $\hat{t}_a$, $\frac{da^*}{dt} = 0$ holds, whereas it is known from the proof of Proposition 2.2 that $\frac{ds^*}{dt}$ is negative. Hence, $\frac{dQ(t)}{dt}$ is negative. It follows from the continuity of the considered functions that there exists a sufficiently small interval $] \hat{t}_Q; \hat{t}_a[$, in which the decrease of $s^*$ outweighs the increase of $a^*$. At the threshold value $\hat{t}_c$, $\frac{da^*}{dt} = 0$, whereas it is known from the proof of Proposition 2.2 that $\frac{da^*}{dt}$ is negative. Hence, $\frac{dQ(t)}{dt}$ is negative. It follows from the continuity of the considered functions that there exists a sufficiently small interval $] \hat{t}_a; \hat{t}_Q[$, in which the decrease in audit effort outweighs the increase in control effort.

QED.

Proof of Proposition 2.4

Because all players except the client firm are kept at their reservation utility of zero, social welfare reads

$$U^{SW} = p(1 - a)Z - (1 - p)(1 - a)(1 - s)tI - \frac{bs^2}{2} - (1 - p)(1 - s)(1 - a)(1 - t)rI - \frac{ka^2}{2}.$$

Partially differentiating $U^{SW}$ w.r.t. $a$ and $s$ and solving the FOCs gives a stationary point at
Now it is necessary to investigate when this stationary point is actually a global maximum of \( U^{SW} \). The Hessian matrix of \( U^{SW} \) reads

\[
H = \begin{bmatrix}
-k & -(1-p)tI - (1-p)(1-t)rI \\
-(1-p)tI - (1-p)(1-t)rI & -b
\end{bmatrix}.
\]

Let the expression \( \frac{\partial^2 U^{SW}}{\partial s \partial a} = -(1-p)tI - (1-p)(1-t)rI \) be denoted by \( A \). Then the characteristic polynomial of the eigenvalue problem \( (H - \lambda I)\vec{x} = 0 \) reads

\[
(-k - \lambda)(-b - \lambda) - (A - \lambda)^2 = 0.
\]

Solving this equation for \( \lambda \) by calculating

\[
(-k - \lambda)(-b - \lambda) - (A - \lambda)^2 = 0 \iff kb + k\lambda + s\lambda + \lambda^2 - (A^2 - 2\lambda A + \lambda^2) = 0 \iff \\
kb + k\lambda + s\lambda - A^2 + 2\lambda A = 0 \iff \lambda = \frac{kb - A^2}{-k - b - 2A} \iff \lambda = -\frac{kb - A^2}{k + b + 2A}.
\]

gives the sole eigenvalue of \( H \). Because \( A < 0 \), it follows that both the numerator and the denominator of the above expression are positive for sufficiently high \( b > \frac{kb}{b} \). Then the eigenvalue \( \lambda \) is negative, which means that \( H \) is negative definite and the stationary point \( \begin{pmatrix} a^{fb} \\ s^{fb} \end{pmatrix} \) is a maximum. Now recall that the equilibrium audit effort \( a^* \) in the auditor-client game reads
\[ a^* = \frac{(1-p)(1-s)(1-t)rI}{k}. \]

Obviously, for \( t = 0 \) the first-best audit effort is

\[ a^{fb} = \frac{(1-p)(1-s^{fb})rI}{k}, \]

which equals the equilibrium effort level \( a^* \) if \( s^* = s^{fb} \) and \( t = 0 \). To check whether this is the case consider the FOC of the client (equation (2.8)) for \( t = 0 \)

\[
\frac{dU^M}{ds} \overset{!}{=} 0 = (1-p)(1-a^*)(1-t)rI - bs \Leftrightarrow \\
\Leftrightarrow s^* = \frac{(1-p)(1-a^*)(1-t)rI}{b},
\]

which in fact equals \( s^{fb} \). Hence, for \( t = 0 \) the first-best solution naturally evolves as the equilibrium out of the game. To show that \( t = 0 \) is the only domain where the first-best solution is an equilibrium, assume that \( a^* = a^{fb} \), which is a necessary condition for a first-best equilibrium. Subtracting the LHS of the FOC (2.8) from the partial derivative \( \frac{\partial U^{SW}}{\partial s} \) gives

\[
\frac{\partial U^{SW}}{\partial s} - \frac{dU^M}{ds} = (1-p)(1-a^{fb})tI - bs + (1-p)(1-a^{fb})(1-t)rI - \left\{ (1-a^*)(1-p)tI - bs + (1-s)(1-p)tI \frac{da^*}{ds} + (1-a^*)(1-p)(1-t)rI \right\} = \\
= -(1-s)(1-p)tI \frac{da^*}{ds},
\]

which is non-zero for \( t > 0 \). Thus \( s^{fb} \) cannot be a solution to (2.8) for \( a^{fb} = a^* \) and \( t > 0 \).

QED.
**Proof of Proposition 2.5**

The full form of equation (2.13) reads

\[
\frac{dU_{SW}^*(a^*, s^*)}{dt} = (1 - p)(1 - s^*)tI \frac{\partial a^*}{\partial t} + (1 - p)(1 - a^*)(1 - s)I(r - 1). \tag{2.19}
\]

The first addend is clearly negative because \( \frac{\partial a^*}{\partial t} \) is negative, while the second addend is clearly positive. Setting (2.19) equal to zero (which is a necessary condition for any \( t \) to be a maximum of \( U_{SW}^*(a^*, s^*) \)) and factoring out common terms gives

\[
(1 - p)(1 - s)I \left[ \frac{\partial a^*}{\partial t} t + (1 - a^*)(r - 1) \right] = 0. \tag{2.20}
\]

Condition (2.20) is satisfied, if either the factor \((1 - p)(1 - s)I\), or the term in brackets becomes zero.

First, consider the case that the term \((1 - p)(1 - s)I\) is zero. It is straightforward that this can only be the case if \( s^* = 1 \) respectively \((1 - s^*) = 0\). The factor \((1 - s^*)\) can be rewritten as

\[
1 - s^* = 1 - \frac{(1 - p)I \left[ k(t + r - rt) - (1 - p)(1 - t)rI(2t + (1 - t)r) \right]}{bk - (1 - p)^2rI^2 \left[ (1 - t)((1 - t)r + 2t) \right]} \Leftrightarrow
\]

\[
1 - s^* = \frac{1}{bk - (1 - p)^2rI^2 \left[ (1 - t)((1 - t)r + 2t) \right]} \left[ bk - (1 - p)^2rI^2 \left[ (1 - t)((1 - t)r + 2t) \right] - (1 - p)I \left[ k(t + r - rt) - (1 - p)(1 - t)rI(2t + r(1 - t)) \right] \right] \Leftrightarrow
\]

\[
\Leftrightarrow 1 - s^* = \frac{k \left[ b - (1 - p)I(t + r - rt) \right]}{bk - (1 - p)^2rI^2 \left[ (1 - t)((1 - t)r + 2t) \right]}.
\tag{2.21}
\]

It is straightforward that this fraction can only become zero if the numerator becomes zero. Setting the numerator of (2.21) equal to zero and solving for \( t \) yields
\[ b - (1 - p)I(t + r - rt) \equiv 0 \Leftrightarrow t^\text{ex}_1 = \frac{-b + (1 - p)rI}{(1 - p)I(r - 1)}. \] (2.22)

Note that the denominator of (2.21) is strictly positive for \( b > \frac{b}{b^5} \). Because the LHS in (2.22) is obviously strictly increasing in \( t \), the numerator of (2.21) and hence the whole expression \( (1 - s^*) \) would become negative for \( t < t^\text{ex}_1 \). Therefore, \( t^\text{ex}_1 \) is the lower bound for \( t \) in the present model. Note that the model only generates meaningful results throughout the whole range of damage apportionment rules \( t \in [0; 1] \) if \( t^\text{ex}_1 \leq 0 \), because otherwise some low values of \( t \) would lead to values of \( s^* > 1 \). Another equilibrium characteristic that can be inferred from the fact that \( (1 - s^*) \) has a unique zero, is that \( s^* \) asymptotically approaches 1, as \( t \) increases. Formally (and purely hypothetically), this means \( \lim_{t \to \infty} \frac{ds^*}{dt} = 0 \).

Now consider the case, that the second part of (2.20) is zero. Formally, this means

\[ \frac{(1 - p)(1 - s^*)rI}{k} t + a^*(r - 1) \equiv r - 1 \Leftrightarrow \]
\[ \frac{(1 - p)(1 - s^*)rI}{k} \frac{t}{k} + (1 - p)(1 - s^*)(1 - t)rI \equiv r - 1 \Leftrightarrow \]
\[ \frac{(1 - p)(1 - s^*)rI}{k} \frac{t + (r - 1)(1 - t)}{k} \equiv r - 1 \Leftrightarrow \]
\[ \frac{(1 - p)(1 - s^*)rI}{k} \frac{t + (r - 1)(1 - t) - (r - 1)}{k} \equiv 0. \]

Inserting the full form of the expression \((1 - s^*)\) from above after the following rearrangements and simplifications gives

\[ b = \frac{(1 - p)^2r(1 - t)^2(r(1 - t) + 2I)}{k}. \] It is obvious that for \( b = b \) the denominator would become zero. Since, \( bk \) is a positive addend, the denominator is greater than zero for every \( b > b \).
Having shown that two domains \( t_1^{ex} \) and \( t_2^{ex} \) that fulfill the necessary condition for an extremum value exist, it remains to prove that one of them is indeed a maximum. We know that \( t_1^{ex} \) is the minimum feasible value of \( t \). For any \( t > t_1^{ex} \), the function can only become zero, if \( t = t_2^{ex} \). Now consider again condition (2.19). If

\[
(1 - p)(1 - s^*) t \frac{\partial a^*}{\partial t} + (1 - p)(1 - a^*)(1 - s^*) I (r - 1) > 0 \Leftrightarrow \]

\[
\begin{align*}
\Leftrightarrow (1 - p)(1 - s^*) I \left[ \frac{\partial a^*}{\partial t} t + (1 - a^*)(r - 1) \right] > 0 \Leftrightarrow \quad (2.23) \\
\Leftrightarrow \frac{\partial a^*}{\partial t} t + (1 - a^*)(r - 1) > 0
\end{align*}
\]

holds, the function \( U^{SW} \) is increasing (decreasing) in \( t \). A product can only be non-zero, if both factors are non-zero. The first factor is strictly positive for all \( t > t_1^{ex} \). Rearranging
the last line of (2.23) and inserting the equilibrium expressions of $a^*$ and $\frac{\partial a^*}{\partial t}$ yields

$$\frac{\partial a^*}{\partial t} t + (1 - a^*)(r - 1) > 0 \Leftrightarrow$$

$$\Leftrightarrow \frac{\partial a^*}{\partial t} t - a^*(r - 1) > -(r - 1) \Leftrightarrow$$

$$\Leftrightarrow -\frac{\partial a^*}{\partial t} t + a^*(r - 1) < r - 1 \Leftrightarrow$$

$$\Leftrightarrow \frac{(1 - s^*)(1 - p)rI}{k} \left[ t + (1 - t)(r - 1) \right] < \underbrace{G(t)}_{(>) r - 1}.$$

For $t = 0$, condition (2.24) simplifies to

$$a^*(r - 1) < r - 1,$$

which is obviously always true for $a^* \in [0; 1]$. Hence, for sufficiently small values of $t$, the function $U^{SW}(a^*, s^*)$ is increasing in $t$. Now consider the other extremum case $t = 1$. For $t = 1$, the equilibrium $a^*$ is zero. Hence, condition (2.24) simplifies to

$$-\frac{\partial a^*}{\partial t} < r - 1.$$

Inserting $s^*(t = 1) = \frac{(1-p)I}{\nu}$ into this condition gives
$$\frac{(1 - p)rI^b(1-p)l}{b} \frac{I^b}{k} < r - 1 \Leftrightarrow$$ (>)

$$\Leftrightarrow (1 - p)rI^b(1-p)l \frac{I^b}{b(r-1)} < k.$$ (>)

This means, if $k$ is smaller than the threshold value $\hat{k}$, the function $U^{SW}(a^*, s^*)$ is decreasing at $t = 1$. It follows, that $G(t)$ crosses the positive constant $(r - 1)$ at some $t^* \in [0; 1]$. Because it is known that only two extremum values exist, this maximum is the sole maximum of $U^{SW}(a^*, s^*)$. However, if $k$ is smaller than $\hat{k}$, then $G(t)$ is always smaller than the constant $(r - 1)$, which means that the function $U^{SW}(a^*, s^*)$ is increasing for all $t \in [0; 1]$. Hence, within the domain range $t \in [0; 1]$, $U^{SW}(a^*, s^*)$ reaches its global maximum at $t = 1$.

QED.
2.9 Appendix B

2.9.1 Outcome of the ICS is observable before the auditor is hired

Now consider the more realistic setting, that the client firm is able to observe the outcome of the internal control procedures before hiring the auditor. In that scenario the payment of the audit fee is contingent on the ICS signal. Only if the ICS does not sort out a bad type project, the auditor is hired. The ICS does not report a signal either because the project is actually of the good type, or because it fails to detect a bad type project. Note that the auditor is able to perfectly observe the state of the ICS after the mandatory ICS audit. Hence, she does not need to update her beliefs about the ICS effectiveness after being hired. Figure 2.6 depicts the game tree of the modified setting.

The modified utility function of the client reads

\[ U^M = p(1 - \alpha)Z - (1 - a)(1 - p)(1 - s)tI - \frac{bs^2}{2} - (p + (1 - p)(1 - s))(F + W). \] (2.26)

If hired, the auditor faces almost the same decision problem as in the base model. The only difference is that given that the ICS has not reported a bad project, the relevant (posterior) probability of auditing a bad type project is now \( \frac{(1 - p)(1 - s)}{p + (1 - p)(1 - s)} \). Differentiating the modified utility function of the auditor

\[ U^A = F - \frac{ka^2}{2} - \frac{(1 - p)(1 - s)}{p + (1 - p)(1 - s)}(1 - a)(1 - t)rI \] (2.27)

w.r.t. \( a \), and setting the derivative equal to zero gives the equilibrium audit effort
Figure 2.6: Game tree of the model with observable ICS outcome

\[
\frac{dU^A}{da} = 0 \iff -ka + \frac{(1-p)(1-s)}{p(1-p)(1-s)}(1-t)rI = 0 \iff a^* = \frac{(1-p)(1-s)(1-t)rI}{k(1-s(1-p))}.
\] (2.28)

It is obvious that the equilibrium audit effort is higher than in the base model. This is because the percentage of audits that exhibit a litigation threat is now greater, since there is no longer a possibility that the ICS avoids bad investment even if the auditor fails to detect a bad project. The partial derivative \( \frac{\partial a^*}{\partial s} \) now reads

\[
\frac{\partial a^*}{\partial s} = -\frac{(1-p)p(1-t)rI}{k(1-s(1-p))^2} < 0.
\]

The modified FOC of the client can be obtained by differentiating (2.26) w.r.t. \( s \). This yields
\[
\frac{dU^M}{ds} \uparrow 0 \iff (1 - a^*)(1 - p)tI + (1 - p)(1 - s^*)tI \frac{\partial a^*}{\partial s} - bs^* + (1 - p)(F + W) - (p + (1 - p)(1 - s^*)) \frac{dF}{ds} \uparrow 0,
\]

(2.29)

whereby the fee effect \( \frac{dF}{ds} = \frac{\partial F}{\partial s} \) reads

\[
\frac{\partial F}{\partial s} = \frac{-p(1 - p)(1 - a)(1 - t)rI}{(1 - s(1 - p))^2}.
\]

Compared to the base model, obviously some incentives change. First, there is a new incentive to exert more control effort, because a more effective ICS increases the probability that the audit fee is saved. Second, the auditor exerts more effort than in the base model. This alters the fee as well as the power of the leverage \( \frac{\partial a^*}{\partial s} \). For the ease of notion, let the probability \((p + (1 - p)(1 - s^*)) = (1 - s(1 - p))\) be denoted by \(g\). Condition (2.29) can then be rearranged to

\[
\iff (1 - a^*)(1 - p)tI + (1 - p)(1 - s^*)tI \frac{\partial a^*}{\partial s} - bs^* + (1 - p)(F + W) + \frac{p(1 - p)(1 - a^*)(1 - t)rI}{g} \uparrow 0.
\]

(2.30)

To ensure the concavity of (2.26) and hence the existence of a unique equilibrium \(s^*\) in the modified model, the second derivative \(\frac{d^2U^M}{d^2s}\) must be negative. Formally, this condition reads

\[58\] This incentive is captured by the positive addend \((1 - p)(F + W)\) in the modified FOC (2.29).
\[
\frac{d^2U_M}{d^2s} < 0 \iff -2(1-p)tI \frac{da^*}{ds} + (1-p)(1-s)tI \frac{\partial^2a^*}{\partial^2s} - b - \frac{p(1-p)(1-t)rI}{g} \frac{da^*}{ds}.
\]

Hence, as in the base model, a lower bound \( b' \) has to be imposed on the control cost parameter \( b \) to ensure the existence of a unique equilibrium.

The effect \( \frac{ds^*}{dt} \) can be worked out by implicitly differentiating the client’s FOC (2.30) w.r.t. \( t \). In the general form, the relevant condition reads

\[
\frac{d^2U_M}{dsdt} = 0 \iff \frac{d^2U_M}{ds\partial t} + \frac{d^2U_M}{ds\partial a} \frac{\partial a^*}{\partial t} + \frac{d^2U_M}{d^2s} \frac{ds^*}{dt} = 0,
\]

which can be rearranged to

\[
\frac{ds^*}{dt} = \frac{\frac{d^2U_M}{ds\partial t} - \frac{d^2U_M}{ds\partial a} \frac{\partial a^*}{\partial t}}{-\frac{d^2U_M}{d^2s}}.
\]

Because \( \frac{d^2U_M}{d^2s} \) is strictly negative for \( b > b' \), the numerator of (2.32) determines the sign of \( \frac{ds^*}{dt} \). The full form of the numerator reads

\[
L(t) = (1-p)(1-a^*)I - \frac{(1-p)^2(1-s^*)(1-a^*)rI}{g} - \frac{p(1-p)(1-a)rI}{g} + (1-p)(1-s^*)I \frac{\partial a^*}{\partial s} + (1-p)(1-s^*)tI \frac{\partial^2a^*}{\partial s^2} - (1-p)tI \frac{\partial a^*}{\partial t} + (1-p) \frac{\partial F}{\partial a} \frac{\partial a^*}{\partial t} \bigg|_{a^*} - \frac{p(1-p)(1-t)rI}{g} \frac{\partial a^*}{\partial t} \iff
\]

\[
\iff - \left[ (1-p)tI + \frac{p(1-p)(1-t)rI}{g} \right] \frac{\partial a^*}{\partial t} + \frac{(1-p)(1-s^*)I}{g} \frac{\partial a^*}{\partial s} + \left[ (1-p)(1-s^*)I \frac{\partial^2a^*}{\partial s^2} \right] + \left[ (1-p)(1-s^*)tI \frac{\partial^2a^*}{\partial s\partial t} \right] + \left[ (1-p)(1-a^*)I - \frac{(1-p)^2(1-s^*)(1-a^*)rI}{g} - \frac{p(1-p)(1-a)rI}{g} \right].
\]

\[
(2.33)
\]
Inserting $a^\ast$ and the full expressions of its partial derivatives, (2.33) simplifies to

$$L(t) = -(1 - p)I \left( \frac{(1 - p)(1 - s)rI(1 - s(1 - p) + p)(-r(1 - t) - 2t + 1)}{k(1 - s(1 - p))^2} + r - 1 \right).$$

(2.34)

$L(t)$ and hence the sign of $\frac{ds^\ast}{dt}$ is negative (positive), if

$$L(t) < 0 \Leftrightarrow \frac{(1 - p)(1 - s)rI(1 - s(1 - p) + p)(-r(1 - t) - 2t + 1)}{k(1 - s(1 - p))^2} > 1 - r \Leftrightarrow V'$$

(2.35)

holds. It is straightforward that the sign of $L(t)$ and therefore the sign of $\frac{ds^\ast}{dt}$ changes for $t$ for which (2.35) holds as an equality. Analogously to the proof of Proposition 2.1, it remains to investigate the behavior of $V'$. The total derivative of $V'$ w.r.t. $t$ generally reads

$$\frac{dV'}{dt} = \frac{\partial V'}{\partial t} + \frac{\partial V'}{\partial s} \frac{ds^\ast}{dt}$$

(2.36)

The partial derivative $\frac{\partial V'}{\partial t}$ thereby reads

$$\frac{\partial V'}{\partial t} = \frac{(1 - p)(r - 2)rI(1 - s)(1 - s(1 - p) + p)}{k(1 - s(1 - p))^2},$$

which is obviously negative for $r < 2$. Because the derivative $\frac{ds^\ast}{dt}$ is by definition zero at the intersection point between $V'$ and the constant RHS in inequality (2.35), $V'$ must
have a negative slope in $t$ at any intersection point. It follows from the continuity of the functions that only one intersection domain $\hat{t}_c$ which fulfills this condition can exist. For all $t < \hat{t}_c$, $V'$ is greater than the constant RHS of (2.35), which means that $L(t)$ and thus $\frac{ds^*}{dt}$ is negative. For all $t > \hat{t}_c$, $V'$ is smaller than the constant RHS of (2.35), which means that $L(t)$ and thus $\frac{ds^*}{dt}$ is positive. Because the partial derivative $\frac{\partial V'}{\partial s}$ in (2.36) is negative due to

$$\frac{\partial V'}{\partial s} = \frac{2(1-p)p^2rI(1-r(1-t) - 2t)}{k(1-s(1-p))^3} < 0,$$

the LHS can be shifted arbitrarily, easiest by varying $b$. Hence, it is only a matter of parameter choice to place the threshold value $\hat{t}_c$ within the interval $[0; 1]$. Overall, it shows that the mechanics and the results do not qualitatively change compared to the base model, because at the time the client decides on the ICS quality, the relevant incentives are the same. However, the relative strengths of the four incentives that drive the client’s equilibrium reaction $\frac{ds^*}{dt}$ change. On the one hand, the fee-saving incentive $(IV)$ in equation (2.10) is stronger because the client now has the possibility to fully avoid the costs of the audit by installing an effective ICS.\footnote{This can be easily seen by comparing incentive $(IV)$ in (2.10) with incentive $(IV')$ in (2.33).} On the other hand, incentive $(I)$ in equation (2.10) is stronger as well. Recall that this incentive reflects the client’s incentive to react to the auditor’s direct effort reduction $\frac{\partial a^*}{\partial t}$ by increasing control effort on his own. This incentive is now stronger, because the auditor reacts more sensitively to changes in his liability. Hence the effort reduction the client ex ante takes into account when choosing his equilibrium reaction $\frac{ds^*}{dt}$ is higher than in the base model.\footnote{This can be easily seen by comparing incentive $(Ia)$ in (2.10) with incentive $(I')$ in (2.33).} Furthermore, the strengths of the strategic effect $(II)$ and the positive incentive caused by a weakening of the leverage $(III)$ increase or decrease both at the same rate $\frac{p}{s^*}$.

QED.
2.9.2 Investors receive less than full loss compensation

If the investors do not receive the full investment $I$ from the auditor and the client in case of an audit failure, but only a fraction $D < I$, they set the cost of capital to

$$U^I(\alpha) = p(\alpha Z - I) + (1 - a)(1 - p)(1 - s)(D - I) \leq 0 \Leftrightarrow$$

$$\Leftrightarrow \alpha^* = \frac{-(1 - p)(1 - s)(1 - a)(D - I) + pI}{pZ},$$

which is obviously strictly decreasing in both $s$ and $a$. The client firm’s FOC then reads,

$$\Leftrightarrow -bs + (1 - p)(1 - a^*)tI - \frac{\partial F}{\partial s} + (1 - p)(1 - s^*)tI \frac{da^*}{ds} - \frac{\partial U^M}{\partial \alpha} \left[ \frac{\partial s}{\partial s} + \frac{\partial s^*}{\partial s} \frac{da^*}{ds} \right] = 0.$$

Note that the cost of capital-effect, depicted by the term in brackets, may be either positive or negative, depending on the strength of the leverage $\frac{da^*}{ds}$. The second-order condition reads
Rearranging yields,

\[ \dddot{b} > -2(1 - p) t I \frac{d^2a^*}{ds^2} - \frac{d^2F}{ds^2} - 2 \frac{d^2U^M}{d\alpha^2} \frac{\partial \alpha^*}{\partial a} \frac{da^*}{ds} \]  \hspace{1cm} (2.40)

Note that the first three terms on the RHS of (2.40) are equal to the condition for \( b \), established for the proof of Remark 2.1. Because the term in brackets is strictly negative, the whole last term is negative, which means that the necessary lower bound for the control costs is lower than in the base model.

Because \( \frac{d^2U^M}{ds^2} \) is strictly negative by the assumption \( b > b'' \), analogously to the proof of Proposition 2.1 we have

\[ sgn \left[ \frac{ds^*}{dt} \right] = sgn \left[ \frac{d^2U^M}{dsdt} + \frac{d^2U^M}{dsda} \right] = sgn \left[ N(t) \right], \]  \hspace{1cm} (2.41)

whereby
\[ N(t) = K(t) + \frac{\partial U^M}{\partial \alpha} \left[ \frac{\partial \alpha^*}{\partial t} < 0 + \frac{\partial \alpha^*}{\partial a} \frac{\partial a^*}{\partial t} < 0 \right]. \quad (2.42) \]

Because the product of the three partial derivatives in \( N(t) \) is strictly negative, the ambiguity of \( K(t) \), which determines the sign of \( \frac{dx^*}{dt} \) in the base model, also holds for \( N(t) \). However, because of \( N(t) < K(t) \), the threshold value \( \hat{t}_c \) is higher than in the base model. Note that the impact of the product in (2.42) can be made arbitrarily small by moving \( D \) closer to \( I \). Hence, it is always possible to choose parameters such that the ambiguity of \( K(t) \) has a dominating influence on \( N(t) \).

QED.
3 Audit Quality, Legal Liability, and the Audit Market under Risk Aversion

ABSTRACT: This paper studies how auditors’ risk aversion shapes the audit market and how it interacts with legal liability. The main pillar of the analysis is that the relation between risk aversion and the supplied audit quality is generally non-monotone. Because futile audit costs augment the severity of a liability loss, extending legal liability decreases the audit quality of sufficiently risk-averse auditors in the absence of any strategic interaction. Linking auditors’ supply of audit services with firms’ demand, the model predicts that relaxing the negligence rule aids in breaking up the current Big-N oligopoly, because more risk-averse auditors disproportionately benefit from a reduction in liability risk. However, whether audit standards should be razor-sharp or deliberately vague, depends on the level of investor protection and the cost-efficiency of the audit technology. It turns out that new technologies like blockchain offer a chance for regulators to foster Big-N audit quality and de-concentrate the industry at the same time.

I thank Ralf Ewert, Sebastian Kronenberger, Beatrice Michaeli (Discussant ARW Basel), Henrik Moser (Discussant DART Workshop), Stefan Schantl, Katherine Schipper, Georg Schneider, Marcel Steller, Alfred Wagenhofer, participants of the DART accounting workshop in Graz, session participants at the 40th EAA Annual Congress in Valencia, 2017, session participants at the 79th VHB Annual Meeting in St. Gallen, 2017, session participants at the 10th Accounting Research Workshop in Basel, 2017, and my fellow PhD students Claire Estebanez, Harald Hinterecker, Peter Krenn, Claudia Labbauer, Felix Niggemann, and Elisabeth Plietzsch for their valuable comments.
3.1 Introduction

This paper studies the joint effects of legal liability and auditors’ risk aversion on audit quality and the structure of the audit market. The theoretical auditing literature almost exclusively treats auditors as risk-neutral agents. However, the responsibility for the audit lies with the individual audit partner, and there is indisputable evidence that individuals tend to behave risk-aversely under uncertainty. The conventional wisdom is that higher risk aversion makes auditors more concerned about the issuance of a wrong opinion and thereby provides an additional incentive to supply high quality audits. Given such a monotone relation, explicitly modeling risk aversion should not yield new insights. However, a closer look reveals that things are less clear than intuition suggests.

The contribution of the paper is that it shows how risk aversion influences the audit market structure, and how a regulator should deal with those effects in liability regulation. Very recently two major cases of auditor liability hit the financial press headlines. In July 2018 PricewaterhouseCoopers LLP was sentenced to pay a whopping 625.3 million $ in damages for failing to detect accounting malpractices that led to the failure of Alabama’s Colonial Bank. In February 2018 Deloitte & Touche LLP paid 149.5 $ million to settle allegations in a related case. Such excessive liability payments carry several consequences for the profession. First, they change the expectations about the litigation threat auditors consider when accepting engagements and planning audits. Second, excessive liability might lead to failures or exits of audit companies. Given the extreme concentration of the audit markets in many economies, there are increasing worries, aptly summarized by PCAOB board member Steven B. Harris in his speech at the 2017 International Institute on Audit Regulation:

"Given this level of concentration, investors have raised concerns as to whether the sta-
tus quo results in lower quality audits and reduced investor protection, and whether our financial markets could cope if one of the Big Four were to fail or exit the field. In other words, are the Big Four too big or too few to fail? And can regulators respond effectively to unscrupulous auditor behavior if they fear their actions could potentially lead to less competition in the audit industry?"  

The analysis in this paper speaks directly to this concern. Using a Dye (1995)-style market-model, I first establish a link between risk aversion, audit quality, and the market structure. The centerpiece of the analysis is that beyond a threshold, higher risk aversion implies lower audit quality. Think of the auditor as a driver going through a crossroad in dense fog. By increasing his speed the driver reduces the chance of having an accident because he reduces the time of being exposed to cross-traffic. On the downside, if he has an accident, then it is more severe. In the same way, an auditor who audits more carefully reduces the chance of suffering liability. However, the higher effort costs make an eventual wealth loss from legal liability more severe. Sufficiently risk-averse auditors are more concerned about the latter case. To limit this worst-case, they cut down their effort and save on audit costs. Empirically, one should observe the following: As competition for clients increases, the small and highly risk-averse auditors are forced to exit the market early, because they offer an inferior audit quality at an uncompetitive price. Mid-tier audit firms, who have an intermediate degree of risk aversion offer the highest audit quality, but can only capitalize on this competitive advantage if audit quality is a particularly valuable good due to weak investor protection. However, if investors can easily recover damages from auditors, then large, approximately risk-neutral firms dominate the market.

I then evaluate which kind of liability regulation is suitable to improve the situation by (a) fostering the audit quality supplied by Big-N firms, and (b) reducing industry

---


concentration. The main insights are as follows: First, increasing the absolute liability threat by tightening investor protection forces small firms to reduce their audit quality in order to limit the worst-case loss, while raising the audit quality supplied by large firms. This tends to aggravate audit market concentration. Second, making audit standards as precise as possible opens up the market to smaller auditors, because compliance with the standard completely eliminates the liability risk. On the downside, the small auditors’ gain in competitiveness lives on the negligent behavior of large auditors. Third, under certain conditions, deliberately vague audit standards can simultaneously aid in achieving both goals. This is because under vague standards the liability risk cannot be completely eliminated. Instead, each audit firm chooses how much of the liability risk to eliminate. Given that auditing is sufficiently costly, a regulator can design a vague negligence system such that (i) both Big-N and mid-tier firms supply higher quality audits, but (ii) mid-tier firms gain a competitive advantage because they opt to eliminate more of the liability risk by increasing their audit quality at a higher rate.

The remainder of the paper is organized as follows: Section two briefly discusses the related literature. Section three contains the analysis and the paper’s main results. Section four concludes the paper by discussing policy recommendations. All formal proofs and derivations are to be found in the appendix.

3.2 Related Literature

From a decision-theoretic perspective, the auditor’s decision is a problem of ”self-protection”. Self-protection is a situation in which an agent can exert costly effort to reduce the probability of a loss. This type of decision problem has originally been described by Ehrlich/Becker (1972). Dionne/Eeckhoudt (1985) were the first to show that the effect of risk aversion on self-protecting effort is ambiguous for certain utility specifications, that is a higher degree of risk aversion can make the costly avoidance of a loss more or less
desirable. They conclude that this finding is "counter-intuitive" and goes against the "widely accepted result that increased risk aversion reduces risky activities". Briys / Schlesinger (1990) were the first to show that this result is independent of the decision-maker’s preferences, while Jullien / Salanié / Salanié (1999) demonstrate that it also holds for stochastic losses. From then on authors focused on characterizing the result in terms of necessary and sufficient conditions imposed on preferences and probability distributions (Chiu (2000), Eeckhoudt / Gollier (2005), Eeckhoudt / Huang / Tzeng (2012)). However, none of the papers care about legal liability.

To my knowledge, the only auditing theory paper that explicitly focuses on the effect of risk aversion on audit quality and audit fees is the paper by Pummerer / Steller / Baldauf (2013). Using a standard decision model, the authors provide numerical calculations, which indicate that more risk-averse auditors always exert a higher level of due care than risk-neutral auditors. They conclude that it is preferable for the client-firms to have highly risk-averse auditors on the market. However, an increase of legal liability lowers the average audit quality supplied, because it drives more risk-averse auditors out of the market. In contrary to Pummerer / Steller / Baldauf (2013), I do not rely on numerical calculations but do a general formal analysis, which is why my findings partially contradict theirs.

Two papers that explicitly consider risk-averse auditors are Balachandran / Nagarajan (1987) and Ewert / Feess / Nell (2000). Both papers compare strict liability and the negligence liability system, taking liability insurance into account. Since in the absence of risk aversion no auditor would ever sign an insurance contract, the consideration of risk aversion is necessary for the research question. Other than these, there are only a few papers in which the authors verbally extend their considerations to the case of risk-averse

---

64 Dionne / Eeckhoudt (1985), p 41 et seq.
65 See in detail Pummerer / Steller / Baldauf (2013), p 43 et seq.
Auditors and thereby underline the robustness of their results.\textsuperscript{66} Among these, the paper by Moore / Scott (1989) stands out. The authors analyze how the extent of auditors’ liability influences a risk-neutral auditor’s effort in the equilibrium of an auditor-client game. In the footnotes they formulate the suspicion that under the assumption of risk aversion the auditor might exert more or less effort than under risk-neutrality.\textsuperscript{67} They argue that increasing risk aversion makes the auditor evaluate the liability consequences as well as the direct costs of auditing higher. However, Moore / Scott (1989) do not elaborate further on their suspicion.

The formal model in this paper is strongly based on the seminal work of Dye (1995). Dye (1995) investigates how auditors’ ability to shield their wealth from liability by forming corporations affects the supplied audit quality, the structure of the audit market and client welfare. He shows that both very wealthy and very poor auditors do not participate on the market, because they offer insufficient price-performance ratios. Allowing auditors to incorporate shifts competitiveness towards large firms, but increases aggregate shareholder wealth, since it eliminates excessive auditing. The Dye (1995)-model is powerful because it embeds the individual audit firm’s decision problem into a complete supply-and-demand-framework, which allows to draw empirical predictions and policy recommendations regarding the ”macro”-effects of regulation. In Dye (1995) all auditors are risk-neutral, but differ in terms of their initial wealth, whereas in my model auditors differ in terms of their risk aversion, but not necessarily in their wealth levels. As I will show, risk aversion entails more complex effects on audit quality and the market equilibrium than wealth.\textsuperscript{68}

Furthermore, my paper complements on the existing literature on liability rules. Among

\textsuperscript{66} Examples include the papers by Moore / Scott (1989), Schwartz (1998), Hillegeist (1999), and Radhakrishnan (1999).

\textsuperscript{67} See footnote 10 in Moore / Scott (1989), p 759.

\textsuperscript{68} In the footnotes Dye (1995) extends his main results to the case of risk-averse auditors. However, he does not elaborate on the effects of varying risk aversion, but imposes additional conditions (e.g. constant absolute risk aversion) for the risk-neutral results to go through.
the important papers of this literature sub-strand are Dye (1993), Narayanan (1994), Chan / Pae (1998), Ewert (1999), Radhakrishnan (1999) and Liu / Wang (2006). These papers focus on comparisons of strict liability and variants of negligence liability. Except Liu / Wang (2006), all of them conclude that a negligence system with a vaguely defined level of due care induces higher audit quality than the stricter systems, because it exhibits additional positive incentive effects. However, none of these papers consider risk-averse auditors.

3.3 Analysis

3.3.1 Audit Quality under Risk Aversion

Consider the following nonparametric decision model: An audit firm which consists of equity partners and employees plans the audit of a risky client. The equity partners decide on how much audit effort \( a \in \mathbb{R}^+ \) to spend. Audit effort includes all the time, technical equipment and personnel resources invested during the audit process. Audit effort maps into audit quality \( q(a) \), which equals the probability of a successful audit. Let the auditor’s payoffs of a successful audit and an audit failure, given audit effort \( a \), be denoted by \( X_i(a) \in \{X_H(a), X_L(a)\} \in \mathbb{R} \), whereby \( X_H(a) - X_L(a) = V > 0 \) for all \( a \). The parameter \( V \) depicts the liability loss to the auditor. Let \( q(a) \) and \( 1 - q(a) \) denote the probabilities of a successful audit respectively an audit failure. Let \( q'(a) > 0 \) and \( q''(a) < 0 \) hold for all \( a \in \mathbb{R}^+ \), that is audit effort increases the probability of receiving the high payoff \( X_H(a) \) at a decreasing rate. The audit technology is imperfect in the sense that for any finite amount of effort there remains a stochastic component in the audit outcome. Both payoffs \( X_i(a) \) strictly decrease in audit effort, that is \( X'(a) < 0 \) for all \( a > 0 \). Assume \( X''(a) \leq 0 \) and \( X_H'(a) = X_L'(a) \), that is the marginal direct audit costs are non-decreasing and independent of the audit outcome.

Let \( U \) be a continuous, twice differentiable, strictly monotonously increasing, strictly
concave and time-separable Bernoulli utility function of the auditor. Let \( \gamma \) be the degree of risk aversion of an audit firm’s equity partners. More precisely, \( \gamma \) is a parameter that increases the Arrow-Pratt coefficient of absolute risk aversion (ARA) \(-\frac{U''(X(a))}{U'(X(a))}\) locally for all \( X(a) \in [X_L(a); X_H(a)] \). Throughout the paper I refer to the firm’s equity partners as ”the auditor” (she) and assume that less (more) risk-averse equity partners run larger (smaller) audit firms. This assumption is based on the equilibrium firm formation theory of Kihlstrom / Laffont (1979). Kihlstrom / Laffont (1979) derive a general market equilibrium in which (i) less risk-averse individuals decide to become entrepreneurs (= equity partners), whereas more risk-averse individuals become their employees, and (ii) the labor demand of entrepreneurs (= the audit firm size) is decreasing in the entrepreneur’s degree of risk aversion.\(^{69}\)

The Von Neumann-Morgenstern (VNM) utility of an auditor generally reads

\[
\mathbb{E}(U(X(a))) = q(a)U(X_H(a)) + (1 - q(a))U(X_L(a)). \tag{3.1}
\]

Furthermore, assume \( \frac{d^2\mathbb{E}(U(X(a)))}{da^2} < 0 \ \forall \ a > 0 \), that is any extremum argument \( a^* \) of the function is a unique maximum. I abstract from moral hazard problems between the auditor and the employees. The auditor’s effort choice is unique and given by the FOC:

\[
\frac{d\mathbb{E}(U(X(a^*)))}{da} = 0 \iff q'(a^*)[U(X_H(a^*)) - U(X_L(a^*))] + \left[q(a^*)U'(X_H(a^*)) + (1 - q(a^*))U'(X_L(a^*))\right]X'(a^*) = 0. \tag{3.2}
\]

\(^{69}\) The latter result requires that both the total and the marginal output of labor are monotone in the stochastic output component. This condition is satisfied whenever the stochastic component enters the production output multiplicatively which is assumed to be the case here; see Theorem 3 in Kihlstrom / Laffont (1979), p 729 et seq.
Analyzing how the audit effort level \( a^* \) that solves \((3.2)\) varies with the degree of risk aversion leads to the following theorem, which essentially resembles Proposition 4 in Jullien / Salanié / Salanié (1999).

**Theorem.** An increase in the auditor’s degree of risk aversion \( \gamma \) increases (decreases) the supplied audit quality \( a^* \) if the current detection probability of misstatements \( q(a^*) \) is higher (lower) than a threshold value \( q_0(a^*) \).[70]

While the proof in the appendix appears gnarled, the economic intuition behind the non-monotone audit quality is rather easy to understand. Increasing the degree of risk aversion \( \gamma \) locally increases the curvature of a decision maker’s utility function. However, depending on the chosen utility function also other functional properties change. Regarding the auditor’s decision problem the changes in absolute and marginal (dis-)utility at the two equilibrium outcomes \( X_H(a^*) \) and \( X_L(a^*) \) are of relevance. While the signs of the changes \( \frac{\partial \Delta U}{\partial \gamma} \) and \( \frac{\partial U'(X_i(a^*))}{\partial \gamma} \) are both unclear, they define a threshold liability probability \( q_0(a^*) \).

Below \( q_0(a^*) \) the auditor exerts more effort, thereby reducing the probability of liability. However, exerting more effort increases the severity of liability, because the direct audit costs shifts both payoffs \( X_i(a^*) \) to the left. If the liability risk \( (1 - q(a^*)) \) exceeds \( (1 - q_0(a^*)) \), the auditor already puts high emphasis on the ”worst case”. Then the effort invested is not only likely to be futile, but also do the effort costs starkly augment the utility loss from liability. In response, the auditor cuts down her effort to limit her maximal utility loss. In other words, auditors may be negligent not because they care too little about liability, but because they care too much.

Note that for \( \gamma = 0 \), \( U(X_i(a^*)) = X_i(a^*) \) and thus \((3.2)\) simplifies to

---

[70] Proof: See appendix.
\[ q'(a^*)X_H(a^*) + q(a^*)X_L'(a^*) - q'(a^*)X_L(a^*) + (1 - q(a^*))X_L'(a^*) \iff \]
\[ \iff q'(a^*) \left( X_H(a^*) - X_L(a^*) \right) + X_L'(a^*) \iff \]
\[ \iff q'(a^*) = \frac{-X_L'(a^*)}{V}. \] (3.3)

It follows, that audit effort is initially increasing in risk aversion whenever the liability consequences \( V \) are sufficiently high. Henceforth assume w.l.o.g. that this is the case.

### 3.3.2 The Audit Market under Risk Aversion

In order to investigate the effects of liability regulation, we need a framework that links the structure of the audit market to auditors’ risk aversion. The market model is borrowed from [Dye (1995)] who investigates how the market structure is driven by wealth differences among auditors.\(^{71}\) The salient difference between the two models is that in [Dye (1995)] a higher wealth at risk causes a higher audit quality\(^{72}\), whereas in my model a higher degree of risk aversion has an ambiguous impact on audit quality. However, the fact that preferences are not uniquely defined requires some additional structure imposed on the model.

#### Demand for Auditing

Suppose that initially there is a set of \( H \) identical risk-neutral firms and a set of \( A \) auditors who differ only in terms of their risk aversion. Each client firm has the same investment opportunity at hand, which requires an initial investment of \( I \) and with prior probability \( p \) yields future returns of present value \( Z > I \). The firm has not enough funds and thus needs to raise external capital on a competitive capital market. Investors require a share

\(^{71}\) The framework is originally set up in [Dye (1993)].

\(^{72}\) As I will show, this does not necessarily hold true under risk aversion.
of the project returns, such that in expectations they just break even. However, with probability \((1 - p)\) the investment fails and \(I\) is lost. In this case investors sue the auditor and in expectations receive the damage payments of \(V\). I assume imperfect investor protection, that is \(V < I\), but costless lawsuits.\(^{73}\) Each firm faces the decision whether to hire an auditor and if so, which auditor to hire. For now assume that each auditor can audit exactly one client.

Upon hiring an auditor, the firm agrees to pay the fee \(W\). After that the auditor privately chooses audit effort \(a\) which translates into a detection probability \(d(a)\). The probability \(d(a)\) is strictly increasing and concave in \(a\), and approaches one as \(a \to \infty\). Audit effort is productive in that it decreases the probability of undertaking a bad project from \((1 - p)\) to \((1 - p)(1 - d(a))\), whereas \(d''(a) < 0\) and \(d(\infty) = 1\). Let \(q(a) = (1 - (1 - p)(1 - d(a)))\) be the audit quality. Each firm prefers to hire an auditor if the expected payoff with an audit exceeds the expected payoff without an audit. Formally, this means that

\[
N = \left[ pZ(1 - \frac{pI + (1 - p)(1 - d(a))(I - V)}{pZ}) - W \right] - pZ(1 - \frac{I}{pZ}) \Leftrightarrow
\]

\[
N = (1 - p) \left[ d(a)I + (1 - d(a)V \right] - W \tag{3.4}
\]

is positive. In analogy to Dye (1995) \(N\) is the ”net benefit of the audit”, that is the improvement in capital allocation net of the audit fee. The aggregate demand for audits as a function of \(N\) equals

\[
Dem(N) = \begin{cases} 
H & \text{if } N > 0 \\
0 & \text{if } N < 0 \\
 n \in [0; H] & \text{if } N = 0
\end{cases}
\]

\(^{73}\) The latter assumption is realistic in the litigious U.S. environment, where class action suits and contingent legal fees are common.
Supply of Auditing

As shown by Dye (1995), for identical client firms the equilibrium $N$ is identical across all clients because each of them thinks that he can hire any auditor. It follows that a firm is willing to pay an auditor who supplies an audit of quality $q(a)$ a fee such that (3.4) is satisfied as an equality. The audit fee is paid at the end of the period and therefore does not immediately alter the auditor’s degree of risk aversion through the wealth change.

Subsequently, the auditor privately chooses her effort $a$. Under the Kihlstrom / Laffont (1979) model assumptions a monotone inverse relation between the auditors’ degree of risk aversion and their firm size prevails. Hence, the potential clients can observe the auditors’ degrees of risk aversion and conjecture what audit quality an auditor will supply. Each auditor chooses the effort according to (3.2). It is obvious that for $V < I$ the fee

$$W = (1 - p)[d(a)I + (1 - d(a))V] - N$$

reaches its maximum for $d(\infty) = 1$. Then $W_{max} = (1 - p)I - N$.

Each auditor’s second-best alternative to being on the market is to become an employee of an audit firm at a normalized fixed wage of zero. The auditor accepts the engagement if in expectations she receives at least her reservation utility $U = U(0)$. Hereafter I make the following assumptions:

A.1: The probability $p$ is sufficiently low, such that $W_{max} - V < 0$ and thus $X_L(a) < 0$ for all $a$.

A.2: $E(U(X(0))) < U$.

A.3: There exists at least one auditor $j$ who chooses $a^*$, such that $N > 0$ and $E(U(X(a^*_j))) > U$.

A.1 implies that no auditor is paid a fee so large that regardless of the outcome she earns

---

a positive payoff from conducting the audit. A.2 implies that an auditor who supplies $q(0)$ is never hired. A.3 excludes the degenerate cases in which no client is willing to hire an auditor, or no auditor is willing to accept the engagement. Moreover, I impose the following two assumptions on auditors’ preferences:

**A.4:** $U$ is identical for all auditors.

**A.5:** $U'(0)$ is identical for all auditors.

Both assumptions are necessary normalizations to make the preference structures of different auditors comparable.

Finally, assume:

**A.6:** $U$ and $d(a)$ are such, that the equilibrium audit effort $a^*$ has a global maximum at some $\gamma_0 > 0$.

**A.7:** $U$ and $d(a)$ are such, that the equilibrium audit effort $a^*$ is zero at an upper bound $\gamma_{\text{max}}$.

A.6 and A.7 capture the concern expressed in the introductory quote. Being aware that regulators cannot afford to let yet another large audit firm exit the market, Big-N audit firms have developed a certain degree of frivolity towards the legal liability risk. This is why they supply audits of worse quality than mid-tier firms. Small auditors on the other hand are overly concerned about liability, which also makes them supply audits of lower quality. The following Lemma 3.1 immediately follows from A.6 and A.7:

---

75 Higher risk aversion only means that the ARA-coefficient increases. However, a higher ARA-coefficient can come with a higher or a lower function value and slope at $X = 0$. A.4 and A.5 control for those superimposing effects. Both assumptions can be slightly relaxed without qualitatively changing the results.

76 The recent working paper by Pittman / Stein / Valentine (2019) studies the association between audit engagement partners’ risk aversion, as measured by the results of criminal background checks, and audit quality for a U.S. sample. While there seems to be a strong negative association between auditors’ risk aversion and the client’s misreporting propensity, they find no association between auditors’ risk aversion and discretionary accruals, and only a weak association between risk aversion and audit fees. Taking into account a likely selection bias (more risk-averse auditors avoiding clients who have a propensity to misstate, but accepting clients who engage in less severe forms of within-GAAP earnings management), this is in line with the conjecture that the overall relation between auditors’ risk aversion and audit quality is non-monotone rather than linear.
Lemma 3.1. There exists at least one threshold degree of risk aversion \( \gamma_n \in [\gamma_0; \gamma_{\text{max}} ] \), for which a risk-averse auditor exerts the same audit effort as the risk-neutral auditor. For all degrees of risk aversion below any \( \gamma_n \) there are always at least two auditors with different degrees of risk aversion who supply the identical audit quality.

Since the optimal audit effort has a maximum at \( \gamma_0 \) and approaches zero as \( \gamma \to \gamma_{\text{max}} \), it follows that there must be a positive degree of risk aversion, denoted by \( \gamma_n \), where \( a^* \) equals the audit effort of an auditor with \( \gamma = 0 \). For all degrees of risk aversion below this intersection argument, the risk-averse auditors’ efforts exceed the risk-neutral auditor’s effort, whereas for all degrees of risk aversion above \( \gamma_n \) the opposite is true.

We are now ready to investigate an auditor’s market participation decision. The total expected surplus generated by an audit is

\[
S = N + \left[ \mathbb{E}(U(X(a^*))) - U \right],
\]

whereas \( R = \mathbb{E}(U(X(a^*))) - U \) is the auditor’s expected rent, and \( N \) is the client firm’s expected rent. Any auditor for whom \( R \) is positive participates on the market. Analyzing how \( S \) varies in \( \gamma \) gives the following lemma:

Lemma 3.2. (i) For imperfect investor protection \( I - V > 0 \), higher audit quality strictly increases \( S \).

(ii) For sufficiently large \( \gamma \), \( S \) decreases in the degree of the auditor’s risk aversion \( \gamma \).

(iii) If investor protection is sufficiently weak (i.e. \( I - V \) is sufficiently large), then \( S \) reaches its global maximum \( S_{\text{max}} \) at \( \gamma_{\text{opt}} < \gamma_0 \).

(iv) If investor protection is sufficiently strong (i.e. \( I - V \) is sufficiently large), then \( S \) reaches its global maximum \( S_{\text{max}} \) at \( \gamma = 0 \).\(^{77}\)

\(^{77}\) Proof: See appendix.
From a social welfare perspective the auditor always under-invests in effort, because she only considers the transfer payment $V$ when choosing her audit, but should be concerned about minimizing the loss of $I$. Hence, implementing a higher audit quality of hired auditors is socially beneficial and should be the goal of a regulator\footnote{Dye (1995) claims that excessive audit quality (caused by high auditor wealth under unlimited liability) reduces social welfare (see Dye (1995), p 86 et seq.). This happens in the case $V > I$, which I do not consider.} The weaker the investor protection within an economy is (i.e. the greater the differential $(I - V) > 0$ is), the more beneficial it is to have higher quality audits. However, the impact of higher risk aversion is ambiguous and given by

\[
\frac{dS}{d\gamma} = \frac{\partial S}{\partial \gamma} < 0 + \frac{\partial S}{\partial a^*} d\gamma > 0.
\]

(3.5)

Under A.1-A.5, risk aversion has a negative effect on an auditor’s expected rent, because it reduces the expected utility the auditor draws from an uncertain expected payoff. However, if audit quality increases in $\gamma$ and investor protection is sufficiently weak, then the indirect positive effect of higher audit quality outweighs the negative direct impact of risk aversion and the total surplus increases until it reaches a global maximum $S^{max}$ at $\gamma^{opt}$. However, if investor protection is sufficiently strong, then the indirect effect is never strong enough to outweigh the negative direct effect and $S$ declines in the degree of the auditor’s risk aversion for all $\gamma$. Because $\mathbb{E}(U(X(0))) < U$, an auditor with degree of risk aversion $\gamma^{max}$ never accepts the engagement. Together with the fact that $S$ is bounded above established in Lemma 3.2, this implies that for a given $N$ there exist a finite number $n$ intervals $[\gamma^{sup}_i, \gamma^{sup}_{i+1}]$ in which auditors receive $S - N \geq U$ and are thus willing to participate on the market. The boundary values $\gamma^{sup}_i$ and $\gamma^{sup}_{i+1}$ are the intersection points of the $S$-curve with the constant $N + U$. By definition of the threshold values, auditors with degrees of risk aversion $\gamma^{sup}_i$ and $\gamma^{sup}_{i+1}$ earn zero rents, whereas auditors
with $\gamma \in ]\gamma^{\text{sup}}; \bar{\gamma}^{\text{sup}}[$ earn positive rents. Denote the cumulative distribution function of auditors by $F(\gamma)$. Then the supply of audits as a function of $N$ can be summarized by

$$\text{Sup}(N) = \begin{cases} A \times \sum_{i=1}^{N} (F(\gamma^{\text{sup}}) - F(\gamma^{\text{sup}})) & \text{if } N \in [0; S^{\text{max}} - U] \\ 0 & \text{if } N \geq S^{\text{max}} - U \end{cases}.$$ 

The boundedness of $S$ in $\gamma$ implies that $\text{Sup}(N)$ strictly decreases in $N$, since each $\gamma^{\text{sup}}_i$ strictly decreases in $N$, and each $\gamma^{\text{sup}}_i$ weakly increases in $N$. It is also clear that if $(I - V)$ is sufficiently small such that $S$ decreases in $\gamma$ for all $\gamma$ (case (iv) of Lemma 3.2), then the sole $\gamma^{\text{sup}}_i$ equals zero.

**Market Equilibrium**

In equilibrium, the market is cleared. An equilibrium is characterized by (i) a common equilibrium net benefit of an audit $N^*$, (ii) a supply of audits $\text{Sup}(N^*)$, and (iii) a demand for audits $\text{Dem}(N^*)$. The following three types of equilibria are possible:

$$N^* \in ]0; S^{\text{max}} - U[ \text{ and } \text{Sup}(N^*) = \text{Dem}(N^*) \quad (3.6)$$

$$N^* = 0 \text{ and } \text{Sup}(N^*) \leq \text{Dem}(N^*) \quad (3.7)$$

$$N^* = S^{\text{max}} - U \text{ and } \text{Sup}(N^*) \geq \text{Dem}(N^*) \quad (3.8)$$

Because $\text{Sup}(N)$ strictly decreases for all $N$, whereas $\text{Dem}(N)$ is non-decreasing for all $N$, an equilibrium always exists and is unique under the imposed conditions. The equilibrium has the following property:

---

79 The proof is essentially the same as the proof of Theorem 1 in Dye (1995). The only difference is that $S$ is only bounded above, but not necessarily inversely U-shaped. Therefore multiple intervals of auditors who participate on the market may exist, whereas in Dye (1995) there is only one such interval.
Proposition 3.1. Suppose $0 < N^* < S_{\text{max}} - U$.

(i) If investor protection within the economy is sufficiently strong (i.e. $I - V$ is sufficiently small), then the least risk averse auditors on the market earn the highest rents.

(ii) The weaker investor protection within the economy is, the more risk-averse the auditors who earn the highest rents are.

Now suppose that competition increases because some client firms merge or exit the market. Formally this means $Dem(N)$ decreases from $H$ to $H'$ for $N > 0$. Then the number of auditors willing to work exceeds the number of firms willing to hire. As a result, auditors will engage in price competition, that is cut down their fees in order to be preferred over their competitors. When clients have to pay less for their audits, their net benefit of an audit $N$ increases. Conversely, $Sup(N)$ decreases because auditors who already earn low rents cannot afford to cut their fee by much more and thus exit the market (i.e. are soaked up by the large firms). This process continues until $Sup(N^*) = Dem'(N^*) = H'$, whereas $N^* > N^*$. The following proposition gives an important property of this competitive dynamic:

Proposition 3.2. As competition for clients increases, the more risk-averse auditors are always driven out of the market earlier than the competitors who supply the same audit quality.

Lemma 3.1 states that there are small auditors who can keep up with larger, less risk-averse auditors in terms of quality. However, Proposition 3.2 shows that when it comes to competing for engagements, they cannot keep up in terms of the price. The large auditors can always afford to sacrifice enough rents to drive smaller auditors who offer the same audit quality.

The proposition immediately follows from parts (iii) and (iv) of Lemma 3.2.

It could also be that due to technological advancements (i.e. big data computing) the number of audits that each auditor can provide increases, and the same number of audits is provided by $H'$ instead of $H$ auditors.

Proof: See appendix.
quality out of the market. Hence, the only way for a smaller auditor to survive under intense competition is to offer an audit of superior quality.

The following figures illustrate the auditors’ choice of effort $a^*$ (figure 3.1), the distribution of surplus $S$ (figure 3.2) and the market equilibrium (figure 3.3). The parametrization, which satisfies A.1-A.7, is

$$U(X) = -e^{-\gamma(X_i - \frac{1}{2} \ln |\frac{1}{\gamma}|)} - (1 - \frac{1}{\gamma}) + L;$$  \hfill (3.9)

$$d(a) = 1 - e^{-a};$$  \hfill (3.10)

$$k(a) = k_0 a,$$  \hfill (3.11)

whereas $p = 0.75$, $k = 1$, $I = 20$, $V = 5$, $U = 6.5$, and the constant $L = 7.5$. One can see, that a demand shift from $\text{Dem}(N) = H$ to $\text{Dem}'(N) = H'$ leads to an increase in $N$ from $N^*$ to $N^*$, which drives auditors within the interval $[\gamma_{\text{sup}}; \gamma_{\text{sup}}']$ out of the market. The auditors’ aggregate rents decrease from the area (1) + (2) + (3) to (1), whereas the client firms’ aggregate rents increase from zero to the area (3). The points $E$ and $E'$ in figure 3.3 indicate the market equilibria.

Propositions 3.1 and 3.2 can be directly used for empirical studies. If investor protection is strong and the competition for engagements is intense, then audit markets should be most concentrated around Big-N firms. However, the weaker the level of investor protection is, respectively the more important high audit quality becomes, the greater the market shares of mid-tier firms should be. Figure 3.2 illustrates the results. The dark-shaded area (3) represents the aggregate client rents. Whenever I talk about de-concentrating the audit industry hereafter, I mean that under competition the aggregate demand for auditing services is met by a large number of small and mid-tier firms instead of a few Big-N firms. In terms of the model this means that the global maximum of $S$ shifts to the
Figure 3.1: Auditors’ effort choice $a^*$ for different degrees of risk aversion $\gamma$

Figure 3.2: Distribution of surplus $S$
right, that is the auditor types who can sacrifice more rents and survive on the market as client competition gets intense, are smaller and more risk-averse.

### 3.3.3 Effects of Liability Regulation

Having established the link between auditors’ risk aversion, audit quality and the audit market structure, we are finally ready to study the joint effects of risk aversion and legal liability, and to give some advice to policymakers. In what follows, two kinds of liability regulation shall be discussed - changing (i) the magnitude of liability consequences, and (ii) the precision of audit standards.

#### Higher Absolute Liability Consequences

An increase in the absolute liability payments is equivalent to an increase in the payoff difference $V = X_H(a^*) - X_L(a^*)$. Differentiating the RHS of (3.2) w.r.t. $V$ yields the following intriguing result:
Proposition 3.3. Higher liability consequences $V$ decrease the supplied audit quality $q(a^*)$ of auditors with a degree of risk aversion greater than a threshold value $\gamma_V$.\(^{83}\)

Proposition 3.3 goes against the conventional notion that in the absence of strategic interaction higher absolute liability consequences strictly raise the supplied audit quality. The traditional argument is that higher liability payments increase the marginal utility of avoiding liability relative to the marginal disutility of the audit costs, which makes the auditor work harder. However, under risk aversion this incentive is accompanied by a countervailing spillover effect between the liability payments and the direct audit costs. Higher legal liability makes auditing more costly at the margin, because conditional on an audit failure the direct audit costs augment the liability loss even more. If the auditor is highly risk-averse, but the marginal productivity of effort $q'(a)$ is rather low (which is the case if $q(a^*)$ is high) compared to the absolute probability of liability, the latter effect dominates. Then it does not pay for the auditor to overcome the increased liability threat by exerting more effort, since this would render an eventual liability loss substantially more severe, but at the same time would only slightly reduce the probability of being held liable.\(^{84}\)

Figure 3.4 illustrates the effect $\frac{da^*}{dV}$. The parametrization, which satisfies A.1-A.7, is

\[\text{ARA} = \frac{U''(X)}{U'(X)}\] strong, while the marginal effect of effort on the liability probability, as given by the slope $q'(a^*)$, is weak.

---

\(^{83}\) Proof: See appendix.

\(^{84}\) Formally, increasing liability payments shifts the payoff from an audit failure $X_L(a^*)$ to the left. On the one hand, this increases the marginal benefit of exerting effort to avoid liability. On the other hand it increases the expected marginal disutility from the direct audit costs, since the slope of the utility function at the domain $X_L(a^*)$ increases, while the slope at the payoff $X_H(a^*)$ remains unchanged. If the auditor is sufficiently risk-averse, then the average slope increase is by the definition $\text{ARA} = \frac{U''(X)}{U'(X)}$, strong, while the marginal effect of effort on the liability probability, as given by the slope $q'(a^*)$, is weak.
Figure 3.4: Expected utility curves for the logarithmic specification with parameters \( y = 0.05, p = 0.9, k = 1.75, V \in \{15, 17, 19\} \)

\[
\mathbb{E}(U(X)) = (p + (1 - p)d(a)) \left\{ \frac{\ln \left[ -k(a) + \frac{1}{\gamma} \right]}{\gamma} - \frac{\ln \left[ \frac{1}{\gamma} \right]}{\gamma} \right\} + \\
+ (1 - p)(1 - d(a)) \left\{ \frac{\ln \left[ -k(a) - V + \frac{1}{\gamma} \right]}{\gamma} - \frac{\ln \left[ \frac{1}{\gamma} \right]}{\gamma} \right\} + \\
\left\{ \frac{\ln \left[ W + \frac{1}{\gamma} \right]}{\gamma} - \frac{\ln \left[ \frac{1}{\gamma} \right]}{\gamma} \right\} + L; \tag{3.12}
\]

\[
d(a) = 1 - e^{-a};
\]

\[
k(a) = ka.
\]

The figure shows that for sufficiently low liability payments the maximum domain of the utility function (which is the optimum audit effort level \( a^* \)) moves upwards, whereas for sufficiently high liability payments it moves back down.\(^{85}\)

In reality, the magnitude of liability consequences \( V \) depends on two factors. Basically it depends on the firm’s investment policy, that is the amount of raised capital \( I \) that ex

\(^{85}\) The chosen logarithmic specification does not allow for explicit analytical derivations of \( a^* \). For the negative exponential utility specification of the numerical example in the previous section, the effect \( \frac{da^*}{dV} \) indeed never emerges. This is due to the fact that each derivative of the exponential function \( e^x \) equals its antiderivative. Use of the exponential function for both \( U \) and \( d(a) \) allows an analytical solution of the model, but also causes the condition for \( \frac{da^*}{dV} > 0 \) to coincide with the second-order condition \( \frac{d^2\mathbb{E}(U(X(a)))}{da^*} < 0 \). It can be easily shown that in general this is not the case.
post may be wasted due to a wrongful audit opinion. Second, it depends on the degree of investor protection \( s \in [0; 1] \), that is the share of the lost investment that in expectations can be recovered from the auditor. While the regulator cannot directly control firms’ investment policy, he can certainly shape a perception of \( s \). Replacing \( V \) by \( s \ast I \) and analyzing how \( S \) varies in the level of investor protection \( s \), yields the following result:

**Proposition 3.4.** Given firms’ investment policy \( I \), increasing the level of investor protection \( s \) increases the rents of large auditors, but decreases the rents of smaller auditors with a degree of risk aversion larger than \( \gamma_V \). Higher investor protection makes those smaller auditors more prone to be driven out of the market, whereas the competitiveness of large auditors increases.\(^ {86} \)

Increasing the level of investor protection has two effects. First, there is a direct negative effect on risk-averse auditors’ rents, because the expected transfer of liability payments from a risk-averse auditor to a risk-neutral client hurts the auditor more than it benefits the client.\(^ {87} \) Second, there is the yet-known indirect effect via the change in audit effort. We know from Lemma 3.2 that higher audit effort is beneficial because it reduces the auditor’s under-investment in audit quality. However, auditors with a sufficiently high degree of risk aversion \( \gamma > \gamma_V \) reduce their audit effort in reaction to a higher liability threat, which makes the under-investment in audit quality more severe. Then the aggregate effect of increasing investor protection on the surplus generated by an audit is negative. As competition increases, the auditors with \( \gamma > \gamma_V \) are driven out of the market earlier. Worryingly, the condition for \( \frac{da}{ds} \) to be negative (condition (3.24) in the proof of Proposition 3.4) suggests that the auditors who suffer the most from the liability increase are mid-tier firms, who are risk-averse but supply audits of high quality. However, those are the auditors who have the potential to outbid the Big-N firms and promote industry

\(^{86}\) Proof: See appendix.

\(^{87}\) Note, that under ubiquitous risk-neutrality the effect is null, because the aggregate monetary effect of transferring liability is null. This is nicely demonstrated by Laux / Newman (2010) who call this the "triangle effect".
de-concentration. In the worst case those auditors are driven out of the market even absent changes in competition. This happens if the surplus $S$ they generate is already so small, that after the increase in liability they cannot provide a positive $N$ to the clients any more. Then higher investor protection blocks profitable investment projects, thereby impeding efficient capital allocation and reducing social welfare. In that light Propositions 3.3 and 3.4 appear quite momentous. If risk-taking entrepreneurs do not have auditors who share their entrepreneurial spirit, a tightening of investor protection backfires.

Before advancing, let the previous results briefly be opposed to the findings of Dye (1995). In Dye (1995) a liability limitation has a monotone negative effect on audit quality. However, this effect only works on large auditors, who reduce their formerly excessive audit quality to the optimal level, enter the market and immediately drive away small auditors who supply audits of lower quality and net value.\footnote{See Theorem 3 in Dye (1995), p 95.} Conversely, in my model the liability limitation has an ambiguous direct effect on both large and small auditors. Because there is no excessive audit quality under imperfect investor protection, the liability reduction shifts competitiveness from large to smaller auditors instead of the other way round. In contrary to Dye (1995), it is the large auditors who already supply their services, whereas the smaller auditors aspire to enter the market. Because smaller auditors can only enter the market at a weak position, they cannot drive large firms out of the market, and it is also not clear whether they are hired at all. If they are not hired, then the liability reduction is indeed a bad idea, as it is still the large auditors who are hired, but the net value of their audits declines. However, if the newcomers are hired, then the economy should benefit from the additional investments made possible by the liability limitation.\footnote{This is true as long as the net benefit from the additional investments undertaken outweighs the loss in the rents of large auditors.}
Precise Auditing Standards

In the next two sections I study how the design of the liability regime affects audit quality and the market equilibrium under risk aversion. Across the globe auditors face negligence liability, that is they are only held liable if an impartial court concludes that the conducted audit procedures were not in accordance with generally accepted auditing standards (GAAS). GAAS can take two forms - they can be either precise or "vague". Precise audit standards allow auditors to exculpate themselves with certainty by complying with the standards. For standardized audit procedures (i.e. auditing bank accounts or trade receivables) audit standards can be considered approximately precise. However, in general audit procedures are highly client-specific and GAAS are generally formulated, more or less imprecise guidelines. To analyze the effects of the different negligence regimes, suppose that the regulator replaces the prevalent negligence rule A by a rule B that is less strict. Following Dye (1995), a rule B is less strict than rule A if the cumulative distribution function of liability for rule A first-order stochastically dominates the cumulative distribution function for rule B. To keep the exposition as clear as possible, I assume that rule A is the strictest vague negligence liability (VNL) rule one can think of, which is the strict liability rule. It follows from continuity of the model that all the subsequent results hold for the transition from a harsh VNL rule to a more mild one.

In the spirit of Dye (1993), I first look at what happens under a negligence system with precise, ”bright-line” audit standards (hereafter labeled ”precise negligence system”). Let the audit quality prescribed by the prevalent standards be denoted by $a_g$. Moreover, denote the direct costs of exerting effort $a$ by $k(a)$. Then an auditor’s expected utility generally reads

$$
E(U_{pnl}) = \begin{cases} 
  a < a_g : & q(a)U(W - k(a)) + (1 - q(a))U(W - k(a) - V) \\
  a \geq a_g : & U(W - k(a)).
\end{cases}
$$

110
Every auditor who complies with the standard $a_g$ faces no liability, whereas every non-compliant auditor faces the original strict liability consequences. It is clear that $a_g$ imposes an upper bound on the supplied audit quality, because exerting more effort causes direct audit costs without further reducing the liability probability. Given the fee $W$, an auditor complies with the standard if

$$U(W - k(a_g)) > q(a^*)U(W - k(a^*)) + (1 - q(a^*))U(W - k(a^*) - V)$$

holds. Next, consider the following definition:

**Definition** An audit standard $a_g$ is said to be implementable if there exists at least one auditor with degree of risk aversion $\gamma \in [0; \gamma_{max}]$ who prefers to comply with the standard $a_g$.

Now we can derive the following proposition:

**Proposition 3.5.** (i) If the auditing standard $a_g$ is implementable and the liability payments are lower than a threshold value $V_g$, then auditors with $\gamma \in [0; \gamma_g]$ violate the standard ($= they supply a^* < a_g$). Auditors with a degree of risk aversion $\gamma$ that lies within a sufficiently small interval above $\gamma_g$ comply with the standard.

(ii) If the auditing standard $a_g$ is implementable, and the liability payments exceed the threshold value $V_g$, then all auditors with degrees of risk aversion $\gamma \in [0; \hat{\gamma}_n]$ who would supply a higher audit quality than the risk-neutral auditor under strict liability comply with the standard. All auditors with a degree of risk aversion within a sufficiently small interval beyond any $\hat{\gamma}_n$ also comply with the standard.

---

90 As long as a lawsuit is costless for the investors, they will always sue the auditor in case of a wrong opinion. However, if a lawsuit is costly, then investors will only sue with a certain probability, whereas auditors randomize between compliance and violation of the standard (see Chan / Pae (1998) and Ewert (1999)). To keep the analysis simple, I leave out this additional strategic layer and assume that there are no legal costs for the plaintiffs.

91 This definition is essentially equivalent to the definition of a "typical" standard in Dye (1993).
(iii) all compliant auditors receive the same fee $W_g$\(^92\).

The key to Proposition 3.5 is that the precise negligence regime allows auditors to shut down the liability risk by complying with the standard. If there is no liability risk, the degree of risk aversion suddenly becomes irrelevant for the effort choice. Therefore, all compliant auditors supply the same audit quality $q(a_g)$ and receive the same fee $W_g$. However, auditors who do not comply with the standard $a_g$ prefer to save on direct audit costs and accept the liability risk. If the liability consequences are moderate, then auditors with a low degree of risk aversion $\gamma < \gamma_g$ prefer to violate the standard because the perceived liability threat is low. However, for auditors with a higher degree of risk aversion the risk of being held liable strongly dampens the expected utility through the risk premium. Therefore, these auditors comply with the standard. If the liability consequences are basically high, then also auditors with a low degree of risk aversion comply with the standard. It then follows that all auditors who are more risk-averse and under strict liability would supply at least the same audit quality as the risk-neutral auditor prefer to comply with the standard, since otherwise they would bear higher effort-costs and suffer from the risk premium.

Proposition 3.5 has a few interesting implications. First, if there are auditors who are negligent, then these are the bigger, less risk-averse auditors. Smaller, more risk-averse auditors however cannot afford to get caught and therefore follow the rules. This result helps to explain why Big-N auditors were involved in huge accounting scandals, such as Enron, WorldCom or Parmalat around the millennium. Second, increasing the liability consequences $V$ up to $V_g$ is beneficial, as it increases the range of auditors who comply with the standard. This is true because higher liability consequences strictly decrease the equilibrium expected utility under non-compliance\(^93\) while leaving the expected utility under compliance unchanged. Hence, low-risk aversion auditors and auditors with degrees

\(^92\) Proof: See appendix.
\(^93\) See the proof of Proposition 3.5 in the appendix.
of risk aversion around $\gamma_n$ become compliant. Third, introducing a precise negligence system suddenly allows all risk-averse auditors to participate on the market, because the payoff $W - k(a_g)$ is a positive constant and $U(W - k(a_g))$ is thus greater than $U$. To the extent that new auditors enter the market and enable profitable investments, social welfare increases. Fourth, making audit standards more precise is an appropriate measure to de-concentrate the market under the following circumstances:

**Proposition 3.6.** If investor protection is sufficiently weak, and the auditing standard is set such that an auditor with $\gamma_g < \gamma^{opt}$ just prefers to comply, then large audit firms are driven out of the market earlier, that is the range of mid-tier and small audit firms who are preferred over large auditors increases.

If investor protection is sufficiently weak, then a few conditions are simultaneously satisfied. First, $\gamma^{opt}$ is positive (see Lemma 3.2). Second, $V$ being sufficiently low ensures that the firm’s benefit from a standard-conform audit is higher than the expected benefit from a sub-standard audit which also includes liability payments. Therefore the fee paid to a compliant auditor always exceeds the fee paid to a non-compliant auditor. Formally, this means that $d(a_g)I > d(a^*)I + (1 - q(a^*))V \forall \gamma$ holds. Third, we know from Proposition 3.5 that if $V$ is low, some low risk aversion auditors do not comply with the standard. Now assume that the audit standard $a_g$ is set such that the first auditor on the market who complies has a degree of risk aversion $\gamma_g < \gamma^{opt}$. Then there exists a second, more risk-averse auditor $B$ who in equilibrium would generate the same expected rent $S$ as a less risk-averse auditor $A$ if she did not comply. However, auditor $B$ must also comply, because in order to generate the same $S$ under non-compliance she would have to exert more costly effort. Because auditor $B$ receives a greater fee but bears lower expected costs under compliance than she would under non-compliance, it follows that some auditors who are slightly more risk-averse than auditor $B$ also comply and enjoy rents that

---

94 Proof: See appendix.
are greater than the rent of the "marginal" non-compliant auditor with a degree of risk aversion just below $\gamma$.

To summarize, making audit standards precise can be a very effective measure to stimulate audit market de-concentration. It is also superior to altering the absolute liability threat, because compliance of auditors leaves the effort choice of non-compliant auditors unchanged, while changing the liability threat moves the efforts of small and large auditors in opposite directions. The flaw with the precise negligence system is that the competitive advantage small auditors can earn from compliance is fueled by the non-compliance of larger auditors, which is the case under weak investor protection. However, if investor protection is strong, then also the large audit firms comply and draw strictly higher rents from conducting audits than the more risk-averse compliant auditors. What we are looking for is thus a liability system that incentivizes all auditors to choose a higher audit quality, but provides stronger incentives to smaller auditors, thereby giving them a competitive advantage. As I will show in the next section, under certain conditions a negligence system with vague auditing standards can be such a cure-all.

**Vague Auditing Standards**

Consider the following modified expected utility function of the auditor:

\[
\mathbb{E}(U_{\text{val}}) = [q(a) + (1 - q(a))n(a)]U(X_H(a)) + (1 - q(a))(1 - n(a))U(X_L(a)).
\]  
\[ (3.13) \]

The function $n(a)$ depicts the probability that the court considers the supplied audit effort to be standard-conform, and acquits her of negligence. Hereafter, $n(a)$ is called the "assessment function". The function is strictly increasing and concave in $a$ ($n'(a) > 0$). This follows from $\frac{\partial U(E(X))}{\partial \gamma}$ being negative (see the proof of Lemma 3.2 in the appendix).
0; \ n''(a) < 0) and asymptotically approaches one as a approaches infinity. Assume that the new probability of avoiding a loss $z(a) = q(a) + (1 - q(a))n(a)$ is more sensitive to audit effort than the sheer audit technology $q(a)$ in the sense that

$$z''(a) - q''(a) < 0 \Leftrightarrow -q''(a)n(a) - 2q'(a)n'(a) + (1 - q(a))n''(a) < 0 \quad (3.14)$$

for each auditor on the market. In economic terms (3.14) means that if the audit effort is already very high, then a slight increase in effort has little impact on the court’s assessment, whereas if the audit effort is currently low, the impact is strong. W.l.o.g. assume that the introduction of $n(a)$ leaves the decision problem structurally unchanged, that is the auditor’s expected utility is still concave in effort ($\frac{d^2E(U(X(a^*)))}{da^2} < 0$) and A.1-A.7 are still satisfied.\footnote{This can be assumed because the introduction of $n(a)$ only shifts the probabilities in the sense that the payoff $X_L(a^*)$ becomes less likely. However, the new technology $z(a) = q(a) + (1 - q(a))n(a)$ does not have fundamentally different properties than $q(a)$, which is why the structure of the decision problem (especially its concavity in $a$) is preserved.}

To distinguish between the two liability systems, let the optimal effort choice under strict and vague negligence liability be denoted by $a_{sl}^*$, respectively $a_{vnl}^*$. The difference in efforts between the two liability systems is given by the difference of the RHSs of the FOCs under strict liability and VNL for the effort level $a_{sl}^*$. Let this difference be denoted by

$$D = \frac{1 - q(a_{sl}^*)}{\Gamma}n(a_{sl}^*)\Gamma X'(a_{sl}^*) - \frac{1 - q(a_{sl}^*)}{\Gamma}n'(a_{sl}^*) + \frac{q(a_{sl}^*)}{\Gamma}n(a_{sl}^*), \quad (3.15)$$

whereas $\Gamma = -\frac{U'(X_H(a^*)) - U'(X_L(a^*))}{\Delta U}$ is the average ARA over the interval $[X_L(a^*); X_H(a^*)]$ and is strictly increasing in $\gamma$.\footnote{Proof: $-\frac{U'(X(a)) - U'(X(a) - \epsilon)}{U(X(a)) - U(X(a) - \epsilon)} \leq -\frac{U''(X(a))}{\Delta U}$, with $\epsilon$ being an infinitely small number, is increasing in $\gamma$ for all $X(a) \in [X_L(a); X_H(a)]$ by the definition of risk aversion. It follows that $\Gamma = -\frac{U'(X_H(a^*)) - U'(X_L(a^*))}{\Delta U}$ is increasing in $\gamma$ as well.} If $D = 0$, then $a_{sl}^*$ solves the auditor’s FOC under the VNL system as well, and the effort choices under the two regimes are identical. However,
if \( D < 0 \), then \( q'(a_{vnl}) < q'(a_{sl}) \) and thus \( a_{vnl} > a_{sl} \).

Obviously, three effects determine the difference in the effort choices across the liability regimes. First, the bad outcome occurs less often given effort \( a \). This pushes back the negative incentive through the high marginal disutility of the direct audit costs and therefore stimulates more audit quality (effect (I)). This effect is more pronounced the higher the degree of risk aversion is. Second, the auditor has an additional incentive to work, because effort \( a \) becomes more productive in the sense that it has a dual impact on both the detection probability and the court’s assessment (effect (II)). This effect is more pronounced if the court’s assessment is more sensitive to changes in audit quality.

Third, there is a negative incentive to cut down effort, because the probability of being held liable is basically lower under the VNL system (effect (III)). This effect is stronger if given \( a \), the court’s assessment is expected to be more favorable (higher \( n(a) \)). One can also see that assumption (3.14) is equivalent to saying that the sum of (II) and (III) strictly increases in \( a \).

To work out the effect of risk aversion itself, the legal change has to be ”calibrated”, that is the assessment function \( n(a) \) has to be designed in a way, such that an auditor with degree of risk aversion \( \gamma_{eq} \) chooses to supply the same audit quality under both regimes \((D = 0)\). This is always possible, because for any \( a_{sl}^* > 0 \) the strict liability system can be reconstructed by setting \( n(a_{sl}^*) \) and \( n'(a_{sl}^*) \) equal to zero. It is also always possible to achieve \( a_{sl}^* = a_{vnl}^* \) for \( \gamma_{eq} = 0 \) and an ”effective” VNL system with \((n(a_{sl}^*) \neq 0, n'(a_{sl}^*) \neq 0)\). From a regulator’s perspective the calibration point is the reference point for the legal change. In reality it is likely that a regulator wants the leading auditors (i.e. Big-N auditors) to supply more audit quality after the legal change than before. However, as Dye (1993) noted early, the Big-N audit firms are ”more active in the standard setting process” and

\[\text{Proof: If } \gamma = 0, \text{ then } \Gamma = 0 \text{ and given } a_{sl}^*, \text{ the set of assessment mechanisms that have the properties } n(a_{sl}^*) = (1 - q(a_{sl}^*)) \text{ and } n'(a_{sl}^*) = q'(a_{sl}^*) \text{ solves } D = 0. \text{ However, for } \gamma_{eq} > 0 \text{ this is only possible, if the sum of effects (I) and (III) in (3.15) is positive.}\]
"generally opposed to increased standards.\footnote{Dye (1993), pp 888 and 901.} Overall, it seems reasonable to assume that the regulator’s and the Big-N auditors’ interests balance each other such that $\gamma_{eq} = 0$. All results go through for a sufficiently slight positive change in the audit quality supplied by the reference auditor $\gamma_{eq}$.

Let us now consider the case in which the regulator designs the assessment mechanism such that $\gamma_{eq} = 0$. Analysis of the difference term \footnote{Proof: See appendix.} (3.15) then yields the following proposition:

**Proposition 3.7.** Given that the strict liability system is replaced by a VNL system, such that the audit quality supplied under the two liability systems is identical at $\gamma = 0$, then within a sufficiently small interval of risk aversion coefficients beyond zero the audit quality supplied under the VNL system is higher (lower) than under the strict liability regime, if the direct audit costs $-X'(a_{sl}^*)$ are sufficiently high (low).

For a calibration of the liability system, the difference between the audit efforts across the two liability systems is given by the change of \footnote{Proof: See appendix.} (3.15) w.r.t. the degree of risk aversion. Basically the VNL system stimulates higher audit effort under risk aversion because liability occurs less often, and thus limiting the maximal loss is less important (effect I). On top there is the effect of a change in the benchmark effort level $a_{sl}^*$, that is the net of a lower absolute litigation probability (effect II), and a higher effect of effort on the marginal litigation probability (effect III). For $\gamma_{eq} = 0$, effect (I) is muted, whereas effects (II) and (III) are exactly balanced. For slightly higher degrees of risk aversion the benchmark effort $a_{sl}^*$ goes up, which by assumption \footnote{Proof: See appendix.} (3.14) discourages high effort relative to the strict liability regime. However, positive risk aversion enables effect (I), and if the marginal audit costs $-X'(a_{sl}^*)$ are sufficiently high, then the positive incentive outweighs the negative net of effects (II) and (III).

Now turn to the effects of vague auditing standards on the equilibrium market structure.
For the moment assume a moderate level of investor protection $I - V$, such that (3.19) is zero for $\gamma = 0$. The following proposition states conditions under which a replacement of strict liability by a VNL regime aids in de-concentrating the audit market:

**Proposition 3.8.** If a strict liability system with medium investor protection is replaced by a VNL system, such that the audit quality supplied under the two liability systems is identical at $\gamma = 0$, (i) the liability payments $V$ are sufficiently high, and (ii) auditing is sufficiently costly at the margin, then the auditor who earns the highest expected rent from being hired has a non-zero degree of risk aversion under the VNL system, but is risk-neutral under strict liability. Then, under the VNL system moderately risk-averse auditors outlast their larger, less risk-averse competitors as competition increases.

A replacement of strict liability by VNL does a few things to the expected rent that a moderately risk-averse auditor draws from an audit. First, the system transition has a direct effect on the liability risk of the auditor. On the one hand, the liability risk decreases because liability occurs less often, that is the probability of receiving the high payoff $X_H(a^*)$ relatively increases. On the other hand, the lower expected liability translate into a lower certain fee payment. Due to the concavity of the utility function, the certain fee reduction increases the differential $U(X_H(a^*)) - U(X_L(a^*))$, thereby increasing the liability risk. Which effect dominates, depends on the weight of the payoff $X_L(a^*)$ in the auditor’s decision. If the liability consequences are sufficiently high, then $q(a^*)$ is high (see (3.3)) and the former effect dominates. As the liability risk declines under the VNL system, so does the risk premium. Consequently, the expected rent drawn from an audit increases as strict liability is replaced by VNL. Second, we know from Proposition 3.7 that, given sufficiently high marginal direct audit costs the positive effect $\frac{da^*}{d\gamma}$ is stronger at $\gamma = 0$ under the VNL system than it is under strict liability. This implies a higher positive effect $\frac{dS}{da} \frac{da^*}{d\gamma}$ under the VNL system, because $\frac{dS}{da}$ is identical under both systems.

101 Proof: See appendix.
Both the lower risk premium and the higher audit effort render the aggregate effect \( \frac{dS}{d\gamma} \) positive at \( \gamma = 0 \), whereas under the assumed moderate investor protection it is zero under the strict liability regime. It follows from continuity, that for slightly higher levels of investor protection, \( \frac{dS}{d\gamma} \) is positive under the VNL system, whereas it is negative under strict liability.

Summarizing Propositions 3.7 and 3.8, the VNL system can enhance the audit quality supplied by large firms and stimulate audit market de-concentration at the same time. The key to this result is that the vagueness of the standard renders the VNL "finer" than the precise negligence regime. Under the precise negligence regime, it is the regulator who clearly defines a standard-conform audit. Auditors can either comply and reduce their liability risk to zero, or violate the standard and face the original strict liability system. However, under the VNL system each auditor chooses the standard \( a_{vnl}^{*} \) she wants to adhere to by herself. A regulator can calibrate the VNL system such that a whole group of auditors supplies a higher audit quality, but the more risk-averse auditors set "their" standard higher than their less risk-averse competitors. Under moderate investor protection, this shifts the competitiveness away from the Big-N towards mid-tier audit firms.

### 3.4 Regulatory Implications and Conclusion

The primary goal of this analysis is to come up with crisp policy recommendations regarding the design of legal liability systems. The quote in the introduction summarizes two major goals of recent auditing regulation. In the short run, regulators want to foster the audit quality supplied by the large auditors on the market in order to keep financial markets stable. In the long run they aim at breaking up the Big-N oligopoly. In conclusion of the paper I shall evaluate the regulatory measures analyzed above with regard to their suitability for achieving these goals.
First, consider changes in the absolute liability threat. In reality, such changes are made by introducing or removing liability caps. Higher expected liability payments raise the audit quality supplied by sufficiently large audit firms. If clients are willing and able to pay for the increased audit quality, financial markets benefit from this measure. However, things look differently when it comes to the goal of audit market de-concentration. If liability payments are raised, especially mid-tier auditors are in jeopardy of being driven out of the market. This is particularly troublesome because those are the auditors who supply the highest quality audits. On top, the wealth loss of legal liability is likely to increase an auditor’s risk aversion. This effect drives mid-tier firms out of the market even though they might survive the monetary loss, because their fee-quality ratio becomes uncompetitive. Given the results in this paper, excessive litigation rulings, such as the 625.3 million penalty for PwC, have to be clearly rejected. Especially on growth markets with high (audit-)risks the introduction of liability caps is recommended to open the market to smaller auditors and pave the way for more profitable investments.

A more subtle way of liability regulation is changing the design of the negligence rule. A regulator can either make auditing standards razor-sharp, or deliberately keep some imprecision. A precise negligence standard has the irresistible charm of eliminating the liability risk for compliant auditors. Hence, it suddenly opens the market to a wide range of medium-sized and small auditors who could not offer a valuable service before. Whether these auditors are hired under competition, however depends on the strength of investor protection. Only if investor protection is strong, the introduction of precise auditing standards shifts the relative competitiveness from large to mid-tier firms. On the downside, the competitive advantage of mid-tier firms requires negligent behavior of large firms.

While a precise negligence regime cannot foster the audit quality supplied by large auditors and promote market de-concentration at the same time, under certain conditions the VNL regime can. These conditions are met in audits that involve cutting-edge technolo-
gies like distributed ledgers, crypto-currencies, or even the use of artificial intelligence. Such developments pose unexpected challenges for audit firms, because they fundamentally change the way financial transactions are executed and recorded. Since routine procedures cannot be used, audits become quite costly. Due to the technological capacities required, the development in the field is primarily pushed by the large audit firms. If the legal environment is sufficiently strict, then keeping auditing standards vague raises the audit quality supplied by the large firms, but at the same time increases the relative competitiveness of mid-tier firms. Regarding these hot topics, one can recommend the U.S. and European regulators not to come up with quick-fixes in the form of precise standards to which auditors either cannot or do not want to adhere. Rather they should let the leading firms come up with their own solutions first (which they will do, since they smell new business opportunities), and then evaluate whether these are appropriate or not. In that sense, Bitcoin and Co. are not only challenges for the auditing profession, but also chances to level the playing field.
3.5 Appendix

Proof of the Theorem

By the implicit function theorem,

\[ \frac{d^2 \mathbb{E}(U(X(a^*)))}{dad\gamma} = \frac{d^2 \mathbb{E}(U(X(a^*)))}{da\partial\gamma} + \frac{d^2 \mathbb{E}(U(X(a^*)))}{d\gamma} \frac{da^*}{d\gamma} = 0 \Leftrightarrow \]

\[ \frac{d^2 \mathbb{E}(U(X(a^*)))}{da^*} \frac{da^*}{d\gamma} = \frac{d^2 \mathbb{E}(U(X(a^*)))}{d\gamma} \]

must hold. Because the auditor’s expected utility function is strictly concave by assumption in order to ensure a unique maximum, the denominator is strictly positive. Hence, the sign of the derivative \( \frac{da^*}{d\gamma} \) follows the sign of the partial derivative \( \frac{d^2 \mathbb{E}(U(X(a^*)))}{da^*} \), which equals the partial derivative of the FOC w.r.t \( \gamma \).

Now rewrite the FOC (3.2) as

\[ q'(a^*) = \frac{[q(a^*)U'(X_H(a^*)) + (1 - q(a^*))U'(X_L(a^*))] X'(a^*)}{U(X_H(a^*)) - U(X_L(a^*))} \delta \delta U. \quad (3.16) \]

It is obvious that the partial derivative of the FOC w.r.t. \( \gamma \) increases, and thus the optimum audit effort \( a^* \) increases (decreases) whenever the fraction on the RHS of (3.16) decreases (increases). Differentiating the RHS w.r.t. the risk aversion coefficient \( \gamma \) yields the condition
\[ q(a^*) [\frac{\partial U'(X_H(a^*))}{\partial \gamma} \Delta U - U'(X_H(a^*)) \frac{\partial \Delta U}{\partial \gamma}] < (>) \]  
\[ (1 - q(a^*)) [U'(X_L(a^*)) \frac{\partial \Delta U}{\partial \gamma} - \frac{\partial U'(X_L(a^*))}{\partial \gamma} \Delta U]. \]  

for the optimum audit effort \(a^*\) to increase (decrease) in the degree of risk aversion \(\gamma\). It follows from the strict concavity of \(U(X(a))\), that

\[ \frac{\partial U'(X_H(a^*))}{\partial \gamma} < \frac{\partial U'(X_L(a^*))}{\partial \gamma} < \Delta U. \]  

holds. Therefore both bracket terms in inequality (3.17) are negative. It follows that \[ \frac{da^*}{d\gamma} > 0 \] at

\[ \frac{q(a^*)}{1 - q(a^*)} > \frac{U'(X_L(a^*)) \frac{\partial \Delta U}{\partial \gamma} - \frac{\partial U'(X_L(a^*))}{\partial \gamma} \Delta U}{\frac{\partial U'(X_H(a^*))}{\partial \gamma} \Delta U - U'(X_H(a^*)) \frac{\partial \Delta U}{\partial \gamma}}. \]  

Let any threshold detection probability \(q(a)\) that satisfies (3.18) as an equality be denoted by \(q_0(a^*)\). Note that the RHS of (3.18) depends on \(a^*\) but not on \(q(a^*)\), whereas the LHS is strictly increasing in \(q(a^*)\). Hence, whenever \(q(a^*) > q_0(a^*)\), the optimum audit effort \(a^*\) increases (decreases) in the degree of risk aversion \(\gamma\). This proves that a higher degree of risk aversion does not necessarily let the auditor exert more audit effort. Because the sign of \(\frac{da^*}{d\gamma}\) only depends on the equilibrium audit quality \(q(a^*)\), it is clear that the ambiguity of \(\frac{da^*}{d\gamma}\) does not hinge on specific assumptions about the auditor’s utility function.

QED.
Proof of Lemma 3.2

The total derivative of $S$ w.r.t. $\gamma$ reads

\[
\frac{dS}{d\gamma} = \frac{\partial S}{\partial \gamma} + \frac{\partial S}{\partial a} da^* d\gamma \iff \frac{dS}{d\gamma} = \frac{\partial E(U(X(a^*)))}{\partial \gamma} + \frac{\partial E(U(X(a^*)))}{\partial a} da^* d\gamma.
\] (3.19)

The first addend represents the effect of higher risk aversion on the expected utility from the uncertain payoff. The expected utility from an uncertain payoff can be rewritten as the utility from a certain payoff with the same expected value minus the utility from the risk premium $X_\eta$, that is

\[
E(U(X)) = U(E(X)) - U(X_\eta).
\] (3.20)

By definition, a risk-averse decision maker’s Bernoulli utility curve is concave. Mathematically, that is equivalent to saying that Jensen’s strict inequality

\[
E(U(X)) < U(E(X)) \quad \forall X
\] (3.21)

holds. Hence, a risk-averse auditor requires a positive risk premium $X_\eta$. It follows that for an auditor to accept the engagement $U(E(X))$ must be greater than $U_0$, which implies $E(X) = q(a)X_H(a) + (1 - q(a))X_L(a) > 0$. Now recall that an increase in the degree of risk aversion $\gamma$ is equal to a local increase in the normalized curvature $-\frac{U''(X)}{U'(X)}$ of the auditor’s Bernoulli utility function within the interval $[X_H(a); X_L(a)]$. Because $U''(0)$ is constant by assumption A.5, the slope of $U'(X)$ must decrease for all $X \in [0; X_H(a)]$ and increase for all $X \in [0; X_L(a)]$. The former implies that $U(X)$ decreases in the degree of risk aversion for all $X \in [0; X_H(a)]$, which establishes $\frac{\partial E(U(X))}{\partial \gamma} < 0$. Moreover, increasing
\[
\frac{U''(X)}{U'(X)} \text{ also means that at any } X \in [X_H(a); X_L(a)], \text{ the difference between } U'(X + \varepsilon) \text{ and } U'(X - \varepsilon), \text{ with } \varepsilon \text{ being an arbitrary small number, increases. This implies that for all } X \in [X_H(a); X_L(a)], \text{ the difference between }
\]

\[
(1 - q(a))U(X - \varepsilon) + q(a)U(X + \varepsilon) < U[(1 - q(a))(X - \varepsilon) + q(a)(X + \varepsilon)]
\]

increases for all \( q(a) \). It follows that the risk premium \( X_\eta \) must strictly increase, as \( \gamma \) increases. This establishes \( \frac{\partial U(X_\eta)}{\partial \gamma} > 0 \) and \( \frac{\partial E(U(X))}{\partial \gamma} < 0 \).

Next, insert \( X_H(a^*) = W - k(a^*) \), \( X_L(a^*) = W - k(a) - V \) and \( W = (1 - p) [d(a^*) + (1 - d(a^*))] - N \) into (3.1), which gives

\[
\mathbb{E}(U(X(a^*))) = q(a^*)U((1 - p) [d(a^*) I + (1 - d(a^*)) V] - N - k(a^*)) +
\]

\[
(1 - q(a^*))U((1 - p) [d(a^*) I + (1 - d(a^*)) V] - N - k(a^*) - V) \iff
\]

\[
\mathbb{E}(U(X(a^*))) = q(a^*)U((1 - p) [d(a^*) I + (1 - d(a^*)) V] - N - k(a^*)) +
\]

\[
(1 - q(a^*))U((1 - p)d(a^*)(I - V) - N - k(a^*) - pV).
\]

Since the auditor optimizes over this function \textit{given the fee} \( W \), the total derivative of (3.22) w.r.t. \( \gamma \) reads

\[
\frac{dS}{d\gamma} = \frac{\partial \mathbb{E}(U(X(a^*)))}{\partial \gamma} < 0 + FQC \bigg|_{\gamma = 0} + \frac{dQ}{d\gamma} \bigg|_{\gamma = 0}.
\]

(3.23)

The partial derivative \( \frac{\partial \mathbb{E}(U(X(a^*)))}{\partial a} \) being positive proves part (i). It is obvious that \( \frac{dS}{d\gamma} \) is negative whenever \( \frac{da}{d\gamma} \) is negative, which proves part (ii). If \( (I - V) \) is sufficiently large,
the total derivative \( \frac{dS}{d\gamma} \) is positive at \( \gamma = 0 \). Then part (ii) implies the existence of a global maximum \( S^{\text{max}} \) at \( \gamma^{\text{opt}} \), whereas (3.23) also implies that \( \gamma^{\text{opt}} < \gamma_0 \). This proves part (iii). Finally, if \((I - V) = 0\), then the second addend in (3.23) is zero and \( \frac{dS}{d\gamma} \) is negative for all \( \gamma \geq 0 \). It follows from continuity that this is also true for sufficiently small \((I - V)\), which proves part (iv).

QED.

**Proof of Proposition 3.2**

Consider an auditor from the interval \([0; \gamma_0]\) and the auditor with \( \gamma > \gamma_0 \) who supplies the same audit quality. Both auditors have the same liability risk, bear the same costs, and by (3.4) receive the same fee. Thus, \( E(X(a^*)) \) is identical. However, from the proof of Lemma 3.2 we know that \( U(E(X)) - U(X_\eta) \) strictly decreases in \( \gamma \). It follows that the more risk-averse auditor generates a lower \( S \), and is thus forced to exit the market earlier as \( Dem(N^*) \) decreases and \( N^* \) increases.

QED.

**Proof of Proposition 3.3**

By the implicit function theorem, in equilibrium we have

\[
\frac{d^2 E(U(X(a^*)))}{dadV} = \frac{d^2 E(U(X(a^*)))}{da \partial V} + \frac{d^2 E(U(X(a^*)))}{da} \frac{da^*}{dV} = 0 \iff \frac{da^*}{dV} = \frac{d^2 E(U(X(a^*)))}{da \partial V} - \frac{d^2 E(U(X(a^*)))}{da}.
\]

It follows from the concavity of the decision problem that the sign of the derivative \( \frac{da^*}{dV} \) equals the sign of the partial derivative \( \frac{d^2 E(U(X(a^*)))}{da \partial V} \).
Increasing (decreasing) the liability payments $V$ while holding the audit effort constant is equivalent to exogenously decreasing (increasing) $X_L(a^*)$, while holding $X_H(a^*)$ constant. Therefore we have

$$sgn\left[\frac{d^2E(U(X(a^*)))}{da^2V}\right] = -sgn \left[-q'(a^*)U'(X_L(a^*)) + (1 - q(a^*))U''(X_L(a^*))X'(a^*)\right],$$

that is whenever

$$-q'(a^*)U'(X_L(a^*)) + (1 - q(a^*))U''(X_L(a^*))X'(a^*) > 0 \iff \left(1 - q(a^*)\right)U''(X_L(a^*))X'(a^*) > q'(a^*)U'(X_L(a^*)) \iff \frac{U''(X_L(a^*))}{U'(X_L(a^*))} < \frac{q'(a^*)}{(1 - q(a^*))X'(a^*)}$$

holds, \(\frac{da^*}{dV}\) is negative. Inserting \(J3.16\) gives the condition

$$-q(a^*)U'(X_L(a^*)) + (1 - q(a^*))U''(X_L(a^*))X'(a^*) > 0 \iff \left(1 - q(a^*)\right)U''(X_L(a^*))X'(a^*) > q'(a^*)U'(X_L(a^*)) \iff \frac{U''(X_L(a^*))}{U'(X_L(a^*))} < \frac{q(a^*)}{(1 - q(a^*))X'(a^*)}$$

for $a^*$ to decrease in $V$. The LHS is the ARA-coefficient at the domain $X_L(a^*)$. Hence, whenever $\gamma$ is sufficiently large, \(\frac{da^*}{dV}\) is negative.

QED.

**Proof of Proposition 3.4**

Inserting $V = s \ast I$ into \(J3.22\) gives

$$-q(a^*)U'(X_L(a^*)) + (1 - q(a^*))U''(X_L(a^*))X'(a^*) > 0 \iff \left(1 - q(a^*)\right)U''(X_L(a^*))X'(a^*) > q'(a^*)U'(X_L(a^*)) \iff \frac{U''(X_L(a^*))}{U'(X_L(a^*))} < \frac{q(a^*)}{(1 - q(a^*))X'(a^*)}$$

for $a^*$ to decrease in $V$. The LHS is the ARA-coefficient at the domain $X_L(a^*)$. Hence, whenever $\gamma$ is sufficiently large, \(\frac{da^*}{dV}\) is negative.

QED.
\[
\mathbb{E}(U(X(a^*))) = q(a^*)U((1 - p)[d(a^*)I + (1 - d(a^*))sI] - N - k(a^*)) +
(1 - q(a^*))U((1 - p)[d(a^*)I + (1 - d(a^*))sI] - N - k(a^*) - sI) \Leftrightarrow
\]
\[
\mathbb{E}(U(X(a^*))) = q(a^*)U((1 - p)[d(a^*)I + (1 - d(a^*))sI] - N - k(a^*)) +
(1 - q(a^*))U((1 - p)d(a^*)I(1 - s) - N - k(a^*) - psI).
\tag{3.25}
\]

Differentiating (3.25) w.r.t. \(s\) yields

\[
\frac{d\mathbb{E}(U(X(a^*)))}{ds} = \frac{\partial \mathbb{E}(U(X(a^*)))}{\partial s} + \frac{\partial \mathbb{E}(U(X(a^*)))}{\partial a} \frac{da^*}{ds} \Leftrightarrow
\]
\[
\Leftrightarrow q(a^*)U'(X_H(a^*))(1 - p)(1 - d(a^*))I - (1 - q(a^*))U'(X_L(a^*)) [p + (1 - p)d(a^*)] I +
\frac{\partial \mathbb{E}(U(X(a^*)))}{\partial a} \frac{da^*}{ds} \Leftrightarrow
\]
\[
\Leftrightarrow q(a^*) (1 - q(a^*)) I \left[ U'(X_H(a^*)) - U'(X_L(a^*)) \right] + \frac{\partial \mathbb{E}(U(X(a^*)))}{\partial a} \frac{da^*}{ds} \leq 0
\]
\[
\text{and } \frac{\partial \mathbb{E}(U(X(a^*)))}{\partial a} \frac{da^*}{ds} > 0.
\tag{3.26}
\]

We know from Lemma (3.2) that \(S\) increases in \(a\) for \(I - V > 0\), respectively \(s < 1\). Under risk neutrality the first addend is zero, since the slope of \(U\) is identical for all \(X(a^*)\). It follows from (3.3) and the continuity of the model, that \(\frac{da^*}{ds}\) is positive for \(\gamma\) within a sufficiently small interval above zero. Hence, the whole effect \(\frac{d\mathbb{E}(U(X(a^*)))}{ds}\) is positive within this interval. However, under risk aversion the first addend of (3.26) is negative. Moreover, for auditors with a degree of risk aversion \(\gamma > \gamma_V\), the effect \(\frac{da^*}{ds}\) is negative. Therefore the aggregate effect \(\frac{d\mathbb{E}(U(X(a^*)))}{ds}\) is negative.

\[\text{QED}\]


Proof of Proposition 3.5

First consider the case that the risk-neutral auditor prefers to violate the standard, that is

\[ U(-k(a_g)) < q(a^*)U(-k(a^*)) + (1 - q(a^*))U(-k(a^*) - V) \]

holds. It is straightforward that an increase in the liability payments \( V \) decreases the RHS\(^{102} \) while leaving the LHS invariant. Hence, only for \( V \) below a threshold \( V_g \) non-compliance is preferable. However, if the audit standard is implementable, (by definition) one auditor with degree of risk aversion \( \gamma_g \) exists, who complies. It follows from continuity that auditors with \( \gamma \) within a sufficiently small interval beyond \( \gamma_g \) also comply with the standard. This proves part (i).

If however, if \( V > V_g \), then the risk-neutral auditor already complies with the standard. Because \( \frac{\partial E(U(a^*))}{\partial \gamma} \) is known to be negative from (3.2), the auditor with degree of risk aversion \( \hat{\gamma}_n \) also complies with the standard, because she bears the same effort costs but suffers from the risk premium. All auditors with degrees of risk aversion \( \gamma < \hat{\gamma}_n \) comply with the standard, because they bear higher effort costs and suffer from the risk premium. It follows from continuity that also auditors within a sufficiently small interval beyond any \( \hat{\gamma}_n \) prefer to comply with the standard, because the increase in the risk premium outweighs the decrease in direct audit costs. This proves part (ii).

Since all compliant auditors exert effort \( a_g \), the fee \( W \) that solves (3.4), and the effort costs \( k(a_g) \) are identical for all compliant auditors. This proves part (iii).

QED.

---

\(^{102}\)Proof: \( \frac{dE(U(X(a^*))))}{dV} = \frac{\partial E(U(X(a^*))))}{\partial V} = -(1 - q(a^*))U'(-k(a^*) - V) < 0. \)
Proof of Proposition 3.6

We know from Lemma (3.2) that if the differential $I - V$ is sufficiently large, then $\gamma^{opt}$ is positive. Given $V < V_g$, large auditors do not comply. If the auditor with $\gamma_g < \gamma^{opt}$ (auditor A) just complies, then the auditor with a higher degree of risk aversion who generates the same rent $S$ under strict liability/ non-compliance (auditor B) also complies. This is true because auditor B must exert more effort than auditor A to generate the same $S$ (see (3.19)). From the proof of Proposition (3.5) we know that any auditor who is more risk-averse and exerts more audit effort than a less risk-averse auditor also complies with the standard. A large differential $I - V$ together with a decent standard $a_g$ also ensure that the condition $d(a_g)I > d(a^*)I + (1 - d(a^*))V$ is satisfied. This condition means that a compliant auditor always receives a greater fee than a non-compliant auditor. However, she bears lower expected costs than under strict liability/ non-compliance (otherwise she would not comply). Hence, her rent under compliance must be higher than under non-compliance/ strict liability. It follows from continuity that auditors within a sufficiently small range beyond auditor B’s degree of risk aversion also comply with the standard and receive rents that are greater than the rent of the ”marginal” non-compliant auditor just below $\gamma_g$.

QED.

Proof of Proposition 3.7

Differentiating (3.13) w.r.t. $a$ yields the FOC
Next, rewrite the auditor’s FOC under the VNL system as

\[
\frac{d\mathbb{E}(U_{\text{vnl}})}{da} = 0 \iff [q'(a_{\text{vnl}}) - q'(a_{\text{vnl}})n(a_{\text{vnl}})] + (1 - q(a_{\text{vnl}}))n'(a_{\text{vnl}})]U(X_H) +
\]
\[
[q(a_{\text{vnl}}) + (1 - q(a_{\text{vnl}}))n(a_{\text{vnl}})]U'(X_H)X'(a_{\text{vnl}}) +
\]
\[
[-(1 - n(a_{\text{vnl}}))q'(a_{\text{vnl}}) - n'(a_{\text{vnl}})(1 - q(a_{\text{vnl}}))]U(X_L) +
\]
\[
(1 - q(a_{\text{vnl}}))(1 - n(a_{\text{vnl}}))U'(X_L)X'(a_{\text{vnl}}) \equiv 0 \iff
\]
\[
q'(a_{\text{vnl}})(1 - n(a_{\text{vnl}})) [U(X_H) - U(X_L)] = -(1 - q(a_{\text{vnl}}))n'(a_{\text{vnl}}) [U(X_H) - U(X_L)] -
\]
\[
\{[q(a_{\text{vnl}}) + (1 - q(a_{\text{vnl}}))n(a_{\text{vnl}})]U'(X_H)X'(a_{\text{vnl}}) +
\]
\[
(1 - q(a_{\text{vnl}}))(1 - n(a_{\text{vnl}}))U'(X_L)X'(a_{\text{vnl}}) \equiv 0 \iff
\]
\[
q'(a_{\text{vnl}}) = \frac{1}{(1 - n(a_{\text{vnl}}))U} \{-[1 - q(a_{\text{vnl}})]n'(a_{\text{vnl}})\Delta U -
\]
\[
\{[q(a_{\text{vnl}}) + (1 - q(a_{\text{vnl}}))n(a_{\text{vnl}})]U'(X_H) + (1 - q(a_{\text{vnl}}))(1 - n(a_{\text{vnl}}))U'(X_L)] X'(a_{\text{vnl}})\} \iff
\]
\[
\iff q'(a_{\text{vnl}}) = -\frac{(1 - q(a_{\text{vnl}}))n'(a_{\text{vnl}})}{(1 - n(a_{\text{vnl}}))U} -
\]
\[
\{[q(a_{\text{vnl}}) + (1 - q(a_{\text{vnl}}))n(a_{\text{vnl}})]U'(X_H) + (1 - q(a_{\text{vnl}}))(1 - n(a_{\text{vnl}}))U'(X_L)] X'(a_{\text{vnl}})\} \iff
\]
\[
\iff q'(a_{\text{vnl}}) = -\frac{[q(a_{\text{vnl}})U'(X_H) + (1 - q(a_{\text{vnl}}))U'(X_L)] X'(a_{\text{vnl}})}{(1 - n(a_{\text{vnl}}))U} -
\]
\[
\frac{(1 - q(a_{\text{vnl}}))n(a_{\text{vnl}})[U'(X_H) - U'(X_L)] X'(a_{\text{vnl}}) + (1 - q(a_{\text{vnl}}))n'(a_{\text{vnl}})}{(1 - n(a_{\text{vnl}}))U}.
\]

(3.27)
\[ q'(a_{vnl}^*) = -\frac{[q(a_{vnl}^*)U'(X_H) + (1 - q(a_{vnl}^*))U'(X_L)]X'(a_{vnl}^*) + (1 - q(a_{vnl}^*))n(a_{vnl}^*)[U'(X_H) - U'(X_L)]X'(a_{vnl}^*)}{(1 - n(a_{vnl}^*))\Delta U} \]

It is obvious that the first addend equals the RHS of (3.2). Inserting \( a_{sl}^* \) for \( a_{vnl}^* \) gives the difference term

\[ D = \frac{(1 - q(a_{sl}^*))n(a_{sl}^*)[U'(X_H) - U'(X_L)]X'(a_{sl}^*)}{\Delta U} - (1 - q(a_{sl}^*))n'(a_{sl}^*) + q'(a_{sl}^*)n(a_{sl}^*) \]

\[ \Leftrightarrow D = (1 - q(a_{sl}^*))n(a_{sl}^*)X'(a_{sl}^*) - (1 - q(a_{sl}^*))n'(a_{sl}^*) + q'(a_{sl}^*)n(a_{sl}^*). \]

Differentiating (3.15) w.r.t. \( \gamma \) gives

\[ \frac{dD}{d\gamma} = n(a_{sl}^*)(1 - q(a_{sl}^*))X'(a_{sl}^*)\frac{d\Gamma}{d\gamma} + \{2q'(a_{sl}^*)n'(a_{sl}^*) - n''(a_{sl}^*)(1 - q(a_{sl}^*)) + q''(a_{sl}^*)n'(a_{sl}^*) + n'(a_{sl}^*)(1 - q(a_{sl}^*))X'(a_{sl}^*)\} \frac{da_{sl}^*}{d\gamma}. \]

If this derivative is negative, then \( D \) is negative within a sufficiently small interval above \( \gamma = 0 \), which is equivalent to \( q'(a_{sl}^*) > q'(a_{vnl}^*) \Rightarrow a_{sl}^* < a_{vnl}^* \). Rearranging (3.29) yields the general condition

\[ 132 \]

\[
-X'(a^*_sl) > \frac{\{2q'(a^*_sl)n'(a^*_sl) - n''(a^*_sl)(1 - q(a^*_sl)) + q''(a^*_sl)n(a^*_sl)\} \frac{da^*_sl}{d\gamma} + \left[n'(a^*_sl)(1 - q(a^*_sl)) - q'(a^*_sl)n(a^*_sl)\right] \frac{da^*_sl}{d\gamma}}{n(a^*_sl)(1 - q(a^*_sl))X''(a^*_sl)\Gamma \frac{da^*_sl}{d\gamma}}
\]
Thus, the effect of higher risk aversion on the utility level drawn from the expected payoff $\frac{\partial U}{\partial \gamma}$ is equal under both liability systems.

The effect $\frac{\partial U(X)}{\partial \gamma}$ captures the negative effect of higher risk aversion on the risk premium. This effect is likely to be different under the VNL system, because the distribution that yields the same expected payoff as the strict liability system changes. On the one hand, the fee declines, because given $a^*_{sl} = a^*_{vnl}$ investors in expectations only receive $(1 - p)(1 - d(a^*_{vnl}))(1 - n(a^*_{vnl}))V$ instead of $(1 - p)(1 - d(a^*_{sl}))(1 - n(a^*_{sl}))V$. On the other hand, the probability mass moves from $X_L(a^*_{vnl})$ to $X_H(a^*_{vnl})$, such that the expected payoff remains the same. Geometrically, the risk premium is given by the difference between the function value of the linear connection line between the two function values $U(X_H(a^*))$ and $U(X_L(a^*))$ and the value of $U(X)$ at the corresponding expected value $E(U(X(a^*)))$. If, holding $a^*$ constant (because they are the same under both systems) both domains $X_L(a^*_{vnl})$ and $X_H(a^*_{vnl})$ are lower than $X_L(a^*_{sl})$ and $X_H(a^*_{sl})$ by the amount $(1 - q(a^*))n(a^*)V$ (= the certain fee reduction), then it must be that the connection line between the function values at the two VNL domains intersects with the connection line between the function values of the two strict liability domains. Because a risk-averse auditor’s Bernoulli utility curve is by definition concave, there can only be one intersection point. For all $E(X(a^*_{sl})) = E(X(a^*_{vnl}))$ above (below) this intersection domain, the connection line under the VNL system passes above (below) the connection line under the strict liability system. Therefore, whenever the expected payoff of the auditor is sufficiently high in the sense that the high payoff $X_H(a^*_{vnl})$ has a sufficiently high probability mass, the transition to the VNL system reduces the risk premium. This is the case if $a^*$ is sufficiently high at $\gamma = 0$, which by (3.3) holds true for sufficiently high $V$. It follows from continuity, that at $\gamma = 0$ a marginal increase in risk aversion causes a weaker increase in the risk premium under the VNL system than under strict liability if the probability mass of $X_H(a^*_{vnl})$ is sufficiently high. Then the effect $\frac{dS}{d\gamma}$ is less negative under the VNL system than under the strict liability system at $\gamma = 0$. 

134
Next, turn to the product $\frac{\partial S}{\partial a_{\text{vnl}}} \frac{da_{\text{vnl}}}{d\gamma}$ in (3.19). Generally, the partial derivative $\frac{\partial S}{\partial a_{\text{vnl}}}$ reads

$$\frac{\partial S}{\partial a_{\text{vnl}}} = \text{FOC} + \{(q(a_{\text{vnl}}^{*}) + (1 - q(a_{\text{vnl}}^{*}))n(a_{\text{vnl}}^{*}))U'(X_H(a_{\text{vnl}}^{*})) +
(1 - q(a_{\text{vnl}}^{*}))(1 - n(a_{\text{vnl}}^{*}))U'(X_L(a_{\text{vnl}}^{*}))\} \times
(1 - p) \left[d'(a_{\text{vnl}}^{*})(I - V) + V(d'(a_{\text{vnl}}^{*})n(a_{\text{vnl}}^{*}) - n'(a_{\text{vnl}}^{*})(1 - d(a_{\text{vnl}}^{*}))\right].$$

We know from the proof of Proposition 3.7 that for $a_{sl} = a_{\text{vnl}}^{*}$ at $\gamma = 0$, the sum $(q'(a_{\text{vnl}}^{*})n(a_{\text{vnl}}^{*}) - n'(a_{\text{vnl}}^{*})(1 - q(a_{\text{vnl}}^{*}))) \iff (1 - p)d'(a_{\text{vnl}}^{*})n(a_{\text{vnl}}^{*}) - n'(a_{\text{vnl}}^{*})(1 - d(a_{\text{vnl}}^{*}))$ is zero. Moreover, at $\gamma = 0$, $U'(X_H(a_{\text{vnl}}^{*})) = U'(X_L(a_{\text{vnl}}^{*}))$, and as under strict liability, the probabilities in the parenthetical term add up to one. Hence, the partial derivative $\frac{\partial S}{\partial a_{\text{vnl}}}$ is identically positive under both regimes. However, we know from Proposition 3.7 that for sufficiently high marginal direct audit costs $-X'(a_{\text{vnl}}^{*})$ the initial increase $\frac{da_{\text{vnl}}}{d\gamma}$ is greater under the VNL system than under strict liability. This implies that the positive product is greater under the VNL system than it is under strict liability.

In summary, for $a_{sl} = a_{\text{vnl}}^{*}$ at $\gamma = 0$, both addends in (3.19) are greater under the VNL system than under strict liability. Given that $I - V$ is such that $\frac{dS}{d\gamma}$ is zero at $\gamma = 0$ under strict liability, $\frac{dS}{d\gamma}$ is positive under the VNL system. It follows from continuity, that within a sufficiently small interval of higher investor protection, $\frac{dS}{d\gamma}$ is negative at $\gamma = 0$ under strict liability, but positive at $\gamma = 0$ under the VNL system. This means that the auditor who ultimately survives as competition becomes maximal has a non-zero degree of risk aversion under the VNL system, but is risk-neutral under strict liability.

QED.
4 Sequential Auditor-Client Interaction under Strategic Effort Complementarity

ABSTRACT: This paper seeks to reconcile the common theoretical acceptance that the relation between firm-side corporate governance investments and external auditing is one of substitution with the ambiguous empirical evidence. Using a compact dynamic game-theoretic model, I trace down conditions under which demand factors - such as auditors’ legal liability, or the existence of agency problems within the firm - increase, respectively decrease the equilibrium corporate governance investments of the firm, and the supplied audit quality. The analysis shows that the driving forces of the equilibrium resource allocation are (i) whether internal governance efforts and external auditing are strategic substitutes or complements, and (ii) how this strategic relation is affected by exogenous demand factors. Compared to prior research, these conditions are not exogenously imposed but endogenously emerge in equilibrium. I then put the theory to the data by running a comparative analysis between German and U.S.-firms, and find supportive evidence.

I thank Ralf Ewert, Petri Ferreira, Beatriz Garcia Osma, Joachim Gassen, Wayne Landsman, Rebecca Reineke (Discussant EIASM Workshop), Dennis Voeller, Juliane Voeller (Discussant DART Workshop), Alfred Wagenhofer, session participants at the 33rd EAA Doctoral Colloquium in Valencia, 2017, participants of the DART Accounting Workshop 2018 in Mannheim, session participants at the 41st EAA Annual Congress in Milan, 2018, session participants at the 13th EIASM Workshop on Accounting and Economics in Paris, 2018, and my department colleagues Harald Hinterecker, Felix Niggemann, Elisabeth Plietzsch, and David Windisch for their helpful comments.
4.1 Introduction

There is a lively ongoing debate, about whether firm-side accounting, assurance and governance activities, and the external audit are complements or substitutes. At first glance, the substitution hypothesis is appealing. If the firm’s accounting is basically less prone to errors and/or the internal control- and oversight-mechanisms are effective, the auditor can rely on those mechanisms to some extent, and less auditing work is necessary to keep the overall audit risk acceptably low. This substitution is the essence of the audit risk model. While theoretical research backs the substitution hypothesis, the empirical evidence is mixed. More recent empirical findings rather suggest that firm-side corporate governance investments and external auditing are complements, that is more sophisticated accounting and more powerful governance mechanisms come hand in hand with a higher audit quality.

A major shortcoming of most empirical research in this field is that one of the two institutional factors - usually the upstream firm-side factor - is treated as an isolated, exogenous regressor. This assumption is problematic, because it ignores that the firm-side investment in accounting and governance is the outcome of strategic auditor-client interaction. As such, the nature of the strategic relation between the two tasks is endogenous. This means that in an equilibrium, the firm’s and the auditor’s activities are both determined by a joint set of factors, that is the factors that primarily drive the auditor’s decision (e.g. auditor liability) are passed through to the firm’s decision via the strategic interaction et vice versa. A conventional regression of Y on X fails to map the data to such effects.

Theory papers tend to focus on the interplay of auditing and the (accounting-related) internal control system. Unlike in the empirical literature, the substitution hypothesis prevails. Nelson / Ronen / White (1988) were among the first to investigate how firm-side accounting effort and external auditing interact. Using a static auditor-client model, they show that in equilibrium more internal control effort of the client triggers less
audit effort. [Pae / Yoo (2001)] consider a more realistic sequential setting, in which the audited firm chooses its internal control quality before the client chooses audit quality, and basically get the same result. They explicitly state that their paper “addresses the owner’s and the auditor’s incentive problems associated with their joint production of information” [103]. Similar results are obtained by [Smith / Tiras / Vichitlekarn (2000)] and [Patterson / Smith (2007)] in fraud settings with sequential auditor-client interaction.

While all these papers provide convincing arguments, the growing discrepancy between theory and empirical evidence calls for answers.

What theoretical papers typically don’t shout out from the rooftops is that the seeming substitution they get as a result, is driven by strategic substitution on the input-side. The importance of distinguishing between different types of substitution/ complementarity is highlighted in the seminal paper of [Bulow / Geanakoplos / Klemperer (1985)]. They define two goods as conventional complements (substitutes) if a more ”aggressive” strategy of one firm, for example a higher quantity in Cournot competition, increases (decreases) the other firm’s total profits. [104] Instead, strategic complements (substitutes) are defined by whether a more aggressive strategy of one firm increases (decreases) the other firm’s marginal profits. In most papers strategic substitution is tacitly implied by the specific modeling of the audit technology. An exception is [Pae / Yoo (2001)], who (due to the general nature of their model) openly state that "the key implication of this assumption in our model [strategic substitution, author’s note] is that, ceteris paribus, when the client has a higher-quality internal control system, it is rational for the auditor to reduce his effort [...]" [105]. Of course, it appears sound that a higher quality firm-accounting reduces the incremental information gain through the audit. However, most people would also agree that well-trained accountants or members of the internal control department can provide organizational support, testing and review assistance to the auditor, thereby

---

103 Pae / Yoo (2001), p 350.
105 Pae / Yoo (2001), p 335.
rendering the audit hours more efficient.\footnote{These tasks are listed by \textcite{wallace1984} to motivate the hypothesis that auditing and internal controls are substitutes in equilibrium. However, they are actually good examples for sources of strategic complementarity, which I show to be a main driver of equilibrium complementarity between the two institutions.}

This paper seeks to disentangle the effects of input-side effort-interaction from the resulting output-side effort allocation between firm-side accounting and external auditing in sequential strategic interaction. More precisely, I investigate how strategic complementarity or strategic substitution between the two institutions endogenously arises in equilibrium, and affects the investments into corporate governance institutions. To this end, I construct a compact model, in which a firm and its auditor are both interested in minimizing material misstatements in the financial reporting on an investment project. The firm chooses the quality of its accounting system first. The auditor observes the accounting quality and then chooses the supplied audit quality. The salient feature of the model is that the firm’s internal governance not only increases the \textit{total} effectiveness of the audit technology, but also may increase its \textit{marginal} effectiveness. In the main part of the analysis I use an exogenous increase in the damage compensation to aggrieved investors as shock effect to investigate the equilibrium effects of the different strategic relationships. In an extension of the base model I also analyze how the existence of an agency problem within the firm affects the equilibrium resource allocation.

The central insight of the analysis is that in the class of sequential auditor-client models the equilibrium allocation between firm-side accounting activities and external auditing is determined by two equilibrium properties of the auditor-client relationship. The first property is the elasticity between the two activity levels, that is either strategic substitution or complementarity between firm-side effort and auditing in the sense of \textcite{bulow1985}. The second characteristic is how this elasticity is affected by changes in the institutional environment, e.g. an increase in auditor liability. I show that there exist equilibria, in which changes in the demand factors of corporate
governance and auditing decrease (increase) both the accounting quality of the firm, and
the audit quality supplied by the external auditor. In these equilibria the overall accuracy
of audited financial statements unambiguously decreases (increases).

I then subject the theory to empirical scrutiny. I take two panels of large German and
U.S. firms. Due to the stricter legal environment and the heavily regulated corporate
governance system, the quality of internal governance appears to be substantially higher
in the U.S. than in Germany. I account for the endogeneity of internal governance with the
use of a 2SLS instrumental variable approach. Consistent with the model predictions, the
empirical association turns out to be substitution in the U.S.-sample, but complementarity
for the German firms.

The contribution of the paper is threefold: First and most important, it contributes to the
mostly empirical debate on whether firm-side governance investments and external audit-
ing are complements or substitutes. It adds to both theoretical and empirical literature
by providing a rigorous and complete analysis of what drives equilibrium substitution or
complementarity, as well as making testable predictions for different institutional settings.
Second, the paper contributes to a growing body of theoretical literature that considers
financial statement accuracy as the result of joint information production between the
firm and its external auditor. Third, it adds to the large literature stream on the eco-
nomic effects of auditors’ third party liability. However, compared to prior literature I do
not look at the incentive effects of some sophisticated liability system, but take one step
back and modify the joint production technology, which is the heart of the auditor-client
relationship.

The remainder of the paper is organized as follows. Section two briefly reviews existent
empirical findings. In section three the model is introduced. In section four I derive
the equilibrium and discuss its properties. Section five contains the comparative static
analysis and the main theoretical results. Section six contains the empirical analysis.
Section seven concludes the paper. All proofs and derivations are in the appendix.
4.2 Empirical Evidence

Due to the limited availability of archival data on internal accounting procedures, controls and governance strength, empirical studies frequently use survey data. One of the early substitution-view papers is Wallace (1984). The author asks the question how the extent of external auditing effort, as measured by the audit fees paid, is related to the auditee’s total expenditures for the internal auditing department. Using a sample of 32 large U.S.-firms for the year 1981, the ordinary least squares (OLS)-regression in Wallace (1984) documents a significantly negative relation. Kaplan (1985) directly investigates how the effectiveness of the internal control system affects the external auditor’s work effort. He conducted a laboratory experiment in which 84 U.S.-auditors were required to provide an assessment of internal control strength, and to accordingly plan the audit hours for a given scenario. The results document a significantly negative association between the number of planned audit hours and the quality of the internal control system.

In a similar experiment, Cohen / Hanno (2000) show an inverse relation between the extent of substantive testing procedures, and the strength of corporate governance and management control philosophy. However, because of their laboratory experiment-settings Kaplan (1985) and Cohen / Hanno (2000) blank out any form of communication and assistance from the internal audit department to the external auditor. Hence, the relation is implicitly constrained to be strategic substitution, which, as my analysis will show, seriously drives the observable equilibrium substitution.

Major empirical support for the substitution hypothesis is brought forward by the study of Felix / Gramling / Maletta (2001). The authors regress audit fees on a variable for the internal audit contribution that is constructed using a survey among practicing U.S.-auditors. This variable is designed to capture both drivers of strategic complementarity (relation between internal and external auditors, extent of assistance) and substitution (inherent error risk, assessment of internal audit quality). Felix / Gramling / Maletta (2001) document a significant negative relation between audit fees and the contribution
of internal control quality to the external audit, thereby suggesting that the strategic substitution dominates. However, the variable for internal audit contribution does not capture the "extent of reliance placed on the overall system of internal controls, as reported by the external auditor"\textsuperscript{107} This factor is included as a separate explanatory variable in the regression model and turns out to be not significant at the 10%-confidence-level. Importantly, Felix / Gramling / Maletta (2001) acknowledge the existence of endogeneity and claim robustness of their findings to estimation as a system of simultaneous equations.

A common problem of the studies above is the small sample size of less than a hundred observations. Using a larger sample of over 1.000 clients\textsuperscript{108} Bedard / Johnstone (2004) document a significant negative association between both corporate governance risk and internal control weaknesses as assessed by the auditor in charge of the engagement, and planned audit hours. However, this association is not that surprising, given the formulation of the hypotheses in terms of compliance with the audit risk-model\textsuperscript{109}

What the aforementioned substitution-view papers have in common is that they build on a supply-side argument. This means that it is the client’s pre-existing accounting and internal control environment that incites the external auditor to supply a certain audit quality. In contrary, the complementarity-view papers rather rely on demand-side arguments. Using survey data of 243 Canadian firms, Anderson / Zéghal (1994) document a strong positive association between internal and external auditing expenditures among the large firms in their sample. However, the association remains insignificant among small firms. Anderson / Zéghal (1994) argue that in the large firms, internal auditing is a complex procedure, and its interplay with external auditing goes beyond simple substitution. They conjecture that the internal auditing costs captures a variety of firm-specific factors that simultaneously increase the demand for internal and external assurance. Goodwin-
Stewart / Kent (2006) investigate whether the existence of an audit committee within the client firm, and the internal audit positively affect the extent of audit procedures. Using audit committee and internal auditing information obtained via a questionnaire from 401 Australian companies, the authors document significantly positive associations between the existence of an external auditing committee, respectively the reliance on the internal audit, and external audit fees as a proxy for audit hours and quality. Goodwin-Stewart / Kent (2006) argue that "audit committees demand a higher quality audit" and that "firms that engage in greater internal monitoring through the use of internal audit also demand higher quality external auditing"[110]. A slightly different argument is brought forward by Knechel / Willekens (2006). Using a small Belgian sample, they document positive associations between external auditing fees, and the set of internal control-/governance mechanisms within the audited firm, as measured by self-constructed risk-management and governance indices. They argue that investing in a specific governance instrument shifts costs and benefits of increased control between individual stakeholders. This creates a demand of other stakeholders for diversification through other control instruments, such as external auditing. However, the direct regression of internal control quality on external auditing expenses yields a significantly negative coefficient on the measure for mandatory internal controls, and an insignificant coefficient on the measure for voluntary internal controls. Knechel / Willekens (2006) argue that under regulation internal controls and external auditing are substitutes because the control costs are generally higher. This argument is supported by the results of my theoretical model. Relying on the arguments of Goodwin-Stewart / Kent (2006) and Knechel / Willekens (2006), Hay / Knechel / Ling (2008) find evidence in support of the complementarity-view within a New Zealandish setting. What differentiates this study from the aforementioned ones is that the authors account for the mutual endogeneity of auditing and internal controls with a two-stage-least-squares (2SLS) instrumental variable model.

In summary, existent empirical research suffers from three major problems. First, only the studies by Felix / Gramling / Maletta (2001) and Hay / Knechel / Ling (2008) account for the endogeneity of the internal governance variable. Second, the sample sizes are often very small, which limits explanatory power. Third, each study is undertaken in a specific institutional environment (country), leaving rather unclear what factors drive the relation. In section six of the paper I attempt to test the general model predictions while avoiding those shortcomings. Specifically, I run a cross-country comparison of large German and U.S. firm panels, using the instrumental variable design of Hay / Knechel / Ling (2008).

4.3 The Model

The model has one period and five stages. It involves three risk-neutral players - an investing firm, an auditor, and capital market investors. The sequence of events and all parameters are common knowledge. A firm seeks new capital to finance a risky project. The project could be the firm itself, the acquisition of another firm, or a R&D project. The project requires an initial investment of $I$. Two types of projects exist on the market. A successful project (type-G) yields future returns with a present value of $Z > I$. However, there are also unsuccessful projects (type-B). If such a project is carried out, it yields future returns with a present value of zero, and the initial investment is lost. The firm wants a type-G project to get financed by external investors. In order to get financing, the firm needs a financial report claiming that the project is of the type-G, and the certification of an independent external auditor that there is no evidence of this assertion being wrong.\footnote{The formal condition for when the investment in an audit pays for the firm is stated in section three.} I assume that ex ante the firm knows the economic prospects of the project, but does not know how to cast this information into GAAP-conform accounting numbers. Since, only two project types, and therefore only two correct accounting
treatments exist, capitalizing/ expensing- or impairment-decisions provide examples. For instance, International Accounting Standard No. 38 (IAS 38) requires firms to capitalize development costs if they fulfill certain criteria regarding the certainty and quantifiability of future economic benefits. Otherwise they have to be expensed immediately.\footnote{IAS 38 is a good example, because the criteria for the recognition of development costs (IAS 38.57) directly relate to the project’s economic prospects. This means that the accounting rule should perfectly separate type-G and type-B projects. Other examples than R&D are impairments of assets, the classification of leases, or the recognition of provisions/ contingent liabilities.}

Prior to the balance-sheet date (or the closing date for an interim report) the firm privately observes a perfectly accurate non-accounting signal $\theta \in \Theta$ about the project’s future prospects, whereas $\Theta$ is the set of all possible states. The signal $\theta$ is information that cannot be communicated to anybody outside the firm. It may include technical details about the project (construction plans, feasibility studies), results of market-analyses, or other ”soft” information, but no financial accounting data. The firm knows that there exists an accounting threshold $\alpha \in \Theta$. The threshold clearly distinguishes type-G and type-B projects, that is all projects with future prospects $\theta > \alpha (\theta < \alpha)$ are of type-G (type-B). However, the firm does not know where the accounting threshold $\alpha$ lies. It only has a belief $\hat{\alpha}$ about the position of the threshold, which represents its knowledge about how the rules and requirements of the accounting standard relate to the non-accounting information $\theta$. I assume that $\hat{\alpha}$ lies below $\theta$, and thus the firm always believes to have a type-G project.\footnote{It is reasonable to assume that the firm has already screened out projects with $\theta < \hat{\alpha}$.}

To attract investor attention and initiate financing, the firm claims to the market that it has a type-G project at hand. However, this claim is not credible, since the firm does not know the true threshold $\alpha$ and thus the translation of the signal $\theta$ into financial statement data. It follows that in the cases $\hat{\alpha} < \theta < \alpha$, the firm’s type-G claim is unintentionally wrong. Hereafter, I assume $\hat{\alpha} < \alpha$, because these are the interesting cases in which overstatements might occur. This means that the firm follows a lax interpretation of the
criteria for a type-G report prescribed in the respective accounting standard.

To lend credence to the type-G claim, the firm needs audited financial statements that confirm this claim. I model the "production" of the audited report in the following stylized way: After observing the signal $\theta$ and releasing the type-G claim, the firm exerts effort $s \in \mathbb{R}^+$ to generate accurate accounting data for the project out of the non-accounting signal $\theta$. The effort $s$ - hereafter labeled "internal effort" - captures all the time, financial and personnel resources spent onto the preparation of elaborate and detailed financial statements. First, it captures the investment in higher-powered valuation and estimation procedures, or more detailed transaction records, as well as improvements in the presentation of financial statements. Second, $s$ reflects investments in a portfolio of internal assurance and governance mechanisms, such as the internal control system (ICS), the internal audit department, or an audit committee. While $s$ might thereby eliminate errors concerning the most other accounting issues, I assume that it cannot correct a wrong assertion about the project type before the audit, because I am interested in analyzing how the internal effort and the external audit effort jointly produce accurate financial statements. There are two justifications for this (technical) assumption: First, as far as $s$ regards the actual preparation of accounting information, the firm insiders might be unaware of how certain project data translates into accounting figures. Hence, it can prepare accounting data to justify the chosen accounting treatment, but more internal effort does not change the firm’s assertion about the relation between the data and the true accounting threshold $\alpha$. Second, cooperation between the internal assurance institutions and the auditor is not only best practice, but required by audit standards. In times of "fast close" it is also reasonable to assume that internal control activities regarding

---

114 In other words, more effort $s$ cannot avoid an error regarding the project’s accounting treatment before the audit. However, it determines how obvious the friction between the prepared accounting data and the (wrongful) type-G report is. Think of $s$ as the detail degree of the appendix in a theoretical research paper. For a reviewer it is easier to verify whether proofs are correct or not, if the author writes down all the intermediate steps. However, an erroneous proof does not become correct just by presenting the calculations in greater detail.

115 Some institutions, such as the audit committee, are by definition tied to the audit.
the financial statements run parallel to the external audit. Formally, this means that
the belief \( \hat{\alpha} \) is invariant in the effort level \( s \). Internal effort is costly, whereas for analytical
simplicity a linear cost function \( c(s) = cs \) with a cost parameter \( c > 0 \) is assumed. After
exerting \( s \), the firm submits a preliminary financial report to an independent auditor,
which claims that the project is of the type-G, and - depending on the effort level \( s \) -
contains more or less detailed accounting information to support this assertion.

The task of the auditor is to lend credence to a type-G report. The auditor may ob-
serve the non-accounting signal \( \theta \), but in contrary to the firm he does not immediately
understand its economic substance. Therefore, he cannot immediately say whether the
financial information the firm has made out of \( \theta \) is a correct accounting representation.
However, the auditor has his own idea \( \hat{\theta} \in \Theta \) of what characteristics clearly qualify a
project for a type-G report. Apart from personal ideas and experiences of the auditor,
accounting standards often contain non-exhaustive lists of requirements that allow for a
certain accounting treatment of economic circumstances. A project with \( \theta = \hat{\theta} \) can thus
be considered as the ideal type-G project. I assume \( \theta < \hat{\theta} \) to capture the riskiness of the
project, whereas \( Pr(\theta > \alpha) = p \), and \( Pr(\theta < \alpha) = (1 - p) \). While the auditor initially
has a limited understanding of \( \theta \), he is an expert in the field of financial accounting,
and thus knows the exact position of the accounting threshold \( \alpha \). Before he accepts the
engagement, he conducts a risk-assessment of the client in order to assess the amount of
necessary resources for the audit. For simplicity, I assume that this assessment is costless
and perfectly informs him about the internal effort \( s \). Because at this stage the veracity
of the type-G claim is unrelated to the internal effort \( s \), the auditor cannot infer anything
about the true signal \( \theta \) from observing \( s \).

The auditor has access to the following audit technology that allows him to interpret the
non-accounting information \( \theta \), and thus to give an opinion on the veracity of the firm’s
preliminary type-G report: After observing \( s \), the auditor exerts audit effort \( a \in \mathbb{R}^+ \),
which translates into an audit of a certain quality \( q(a, s) \). Audit quality strictly increases
in both the level of audit effort and the preceding internal effort of the firm. This modeling captures that the auditor finds it easier to detect a misclassification of a type-B project as type-G project, if the firm’s overall accounting system is more sophisticated, as the friction between the type-G claim and the accounting data is more obvious for a professional. After conducting an audit of quality $q(a, s)$, the auditor receives a signal $\tilde{\theta} \in [\theta; \hat{\theta}]$, which follows the distribution $f(\tilde{\theta} \mid q(a, s))$. The observation of $\tilde{\theta}$ captures that by exerting audit effort the auditor learns about the firm’s business, gathers evidence, and thus gains some understanding about the non-accounting information $\theta$. Audit effort is productive in the sense, that it shifts the distribution $f(\tilde{\theta} \mid q(a, s))$ to the left, that is $\frac{dF(\tilde{\theta} \mid q(a, s))}{ds} > 0 \ \forall \ \tilde{\theta}$ (first-order stochastic dominance). Whenever the auditor receives a signal $\tilde{\theta}$ that is below $\alpha$, he has enough evidence to conclude that the type-G assertion of the firm is wrong, and issues a qualified opinion. By restricting the support of the audit evidence distribution $f(\hat{\theta} \mid q(a, s))$ to the interval $\tilde{\theta} \in [\theta; \hat{\theta}]$, I exclude a type-I error. Specifically, I assume that the auditor’s detection technology has the exponential form $d(a, s) = F(\tilde{\theta} \mid q(a, s) < \alpha) = 1 - e^{-sa}$. This probability function has the convenient property that it is naturally bounded between zero and one. The exponential modeling of the audit technology is also the reason why I do not plainly interpret $s$ as internal control quality, such as Pae / Yoo (2001). If $s$ were the quality of the firm’s ICS, one would have to consider the case that the ICS sorts out bad projects before the actual audit. However it is not possible to disentangle a separate detection probability of the ICS from the function $1 - e^{-sa}$, such as it would be possible in a multiplicative model (e.g.

---

116 Audit quality can be considered as the output of a team-production between the firm and the auditor.
117 The modeling of audit quality as a function of audit effort and internal effort is equivalent to Nelson / Ronen / White (1988) and Pae / Yoo (2001).
118 More formally, $\tilde{\theta}$ is an upward-biased, noisy signal about $\theta$.
119 This seems reasonable, because whenever the auditor confronts the firm with a signal $\tilde{\theta} < \theta$, the firm would have enough evidence to convince the auditor that the true signal is at least $\theta$.
120 An exponential audit technology is also assumed by Newman / Patterson / Smith (2005).
\[ d(a, s) = 1 - (1 - s)(1 - a) \]  

With probability \( 1 - e^{-sa} \), the auditor receives a signal \( \tilde{\theta} \in [\theta, \alpha] \) in the "dangerous" case \( \theta < \alpha \), and then correctly qualifies the opinion. Per unit of \( a \) the auditor incurs constant variable costs of \( k \). The cost parameter \( k \) captures the cost-efficiency of the audit technology, that is a lower \( k \) represents a more experienced auditor, or the availability of higher-powered tools for data analysis. Figure 4.1 illustrates the production of audited financial statements, whereas \( q'(a, s) > q(a, s) \).  

After performing the audit, the auditor issues either an audit report which contains (i) either a qualified or an unqualified opinion on the firms unaudited financial statements, and (ii) a report on the firm-side error risk, that is the level of \( s^* \). If he detects a type-B project (qualified opinion), investors provide no capital and the game ends. However, if he

\[^{121}\text{However, the model results would remain qualitatively unchanged if internal effort } s \text{ could sort out misstatements prior to the audit, because ex ante the auditor would still be hired with a positive probability in equilibrium.}\]

\[^{122}\text{Below the line, the modeling is equivalent to the frequently used scenario that the investing firm does not know the project type, and ignorantly submits a high report. Examples include the papers by Dye (1993), Chan / Pae (1998), Radhakrishnan (1999), Liu / Wang (2006), Lu / Sapra (2009) and Laux / Newman (2010). This assumption allows misreporting by the client without considering active misreporting by fraudulent managers. I consider the argument laid-out to be descriptive of the way unintentional errors enter financial statements.}\]

149
confirms the type-G report (unqualified opinion), the project gets financing and is carried out. At the final stage of the game the true project type is revealed. If the auditor’s opinion turns out to be wrong, investors lose their initial investment. I assume that the firm is then bankrupt, and damages can only be partially recovered by suing the auditor who is a deep pocket-defendant. I also assume that the lawsuit against the auditor is costless and always successful, yielding the investors a damage compensation of \( V < I \).

To ensure that the equilibrium audit effort is positive, I assume w.l.o.g.

\[
k < s(1 - p)V,
\]

that is the costs of auditing are sufficiently small compared to the expected liability consequences. Since litigation is often deemed to be an existential threat for auditors, I consider this (technical) assumption to be realistic.

For his services the auditor receives a non-contingent fee \( W \) by the client, which is fixed upon signing the engagement letter. Since at that time the auditor cannot yet observe the true \( s \), \( W \) is based on the conjecture about internal effort \( s^c \). To ensure the auditor’s participation, the fee must at least cover his total costs, that is

\[
W(s^c) \geq ka(s^c) + (1 - p)(1 - d(a(s^c), s^c))V
\]

must hold. I assume that in equilibrium the auditor’s conjecture about internal effort is true \( (s^c = s^\ast) \). After receiving an unqualified opinion, investors agree to finance the project, and set the cost of capital. The capital market is perfectly competitive, that is investors require a share \( \alpha \in [0; 1] \) of the project’s returns, such that in equilibrium they just break even. The reservation utility of the risk-neutral investor community is also

\[123\] This simplified modeling of the auditor’s third party liability is equivalent to a strict liability system without legal costs for the plaintiffs. The same assumption is chosen by [Hilgeist (1999), Pae / Yoo (2001), Newman / Patterson / Smith (2005), and Laux / Newman (2010)].
normalized to zero. The client firm keeps the remaining share \((1 - \alpha)Z\).

Finally, the model can be summarized by the three players’ expected utility functions, denoted by \(U^C\) (client firm), \(U^I\) (investors), and \(U^A\) (auditor):

\[
U^C = p(1 - \alpha)Z - W(s^c) - c(s),
\]

\[
U^I = \frac{p}{p + (1 - p)(1 - d(a, s))}(\alpha Z - I) - \frac{(1 - p)(1 - d(a, s))}{p + (1 - p)(1 - d(a, s))}(I - V) = 0,
\]

and

\[
U^A = W(s^c) - (1 - p)(1 - d(a, s))V - ka.
\]

### 4.4 The Equilibrium

The subgame-perfect equilibrium of the game is established by backward-induction. As will be shown, \(a^*\) is surjective in \(s\). Therefore, investors are able to infer the equilibrium audit effort from the information on internal effort in the audit report. After receiving an audit report containing an unqualified opinion, they set the cost of capital as follows:

\[
U^I = \frac{p}{p + (1 - p)(1 - d(a^*, s^*))}(a^* Z - I) - \frac{(1 - p)(1 - d(a^*, s^*))}{p + (1 - p)(1 - d(a^*, s^*))}(I - V) = 0 \iff \\
\iff a^* = \frac{e^{-sa}(1 - p)(I - V) + pI}{pZ}.
\]

It is straightforward that the equilibrium share of the investors strictly decreases in both
equilibrium effort levels, the possible future returns $Z$, the probability of success $p$, and the liability payments $V$, whereas it strictly increases in the initial investment $I$.

Next, consider the auditor’s equilibrium behavior. Before deriving an equilibrium audit effort, it is necessary to specify when the client decides to hire the auditor at all. Since the auditor and the investor community both receive their reservation utility in equilibrium, they do not care about whether the auditor is hired or not. The firm however, only decides to hire the auditor and spend effort $s$ on the preparation of project-related accounting information if the audit increases its expected utility. Formally, this means that

$$U^C \geq p(1 - \alpha^*)Z \iff p(1 - e^{-s\alpha}(1 - p)(I - V) + pI)Z - W(s^c) - c(s) \geq p(1 - \alpha)Z$$

(4.5)

holds\textsuperscript{124} whereas

$$\alpha = \frac{(1 - p)(I - V) + pI}{pZ}$$

(4.5)

W.l.o.g. assume that (4.5) is always fulfilled, and the auditor is hired by the client. Solving the auditor’s FOC yields

$$\frac{dU^A}{da} = 0 \iff a^* = \frac{ln\left[\frac{s(1-p)V}{k}\right]}{s}.$$  

(4.6)

The following lemma contains the important property of the equilibrium audit effort:

\textsuperscript{124} Condition (4.5) implies that the project’s ex-ante NPV might be slightly negative, but the project is nevertheless pursued.

\textsuperscript{125} Note that $\alpha$ is the upper bound for $\alpha^*$. It is obvious that $\alpha^* = \alpha$ for $d(a, s) = 0 \iff a = 0$. 

152
Lemma 4.1. The equilibrium audit effort \( a^* \) increases in the internal effort \( s \) if internal effort is currently sufficiently low \( (s < \hat{s}) \), whereas it decreases in the internal effort \( s \) if internal effort is currently sufficiently high \( (s > \hat{s}) \).\(^{126}\)

Lemma 4.1 is vital for the further analysis. Given the firm’s choice of internal effort \( s \), the auditor decides on the audit effort he supplies. Intuitively, in a supply-side framework one might expect that a more accurate financial accounting allows the auditor to save on audit effort, since the risk that the pre-audited financial statements are materially overstated is basically lower.\(^{127}\) This substitution is the central element of the audit risk model. However, there is new facet brought into play by the exponential auditing technology. A higher accounting quality not only shifts the error probability down, but also makes the audit technology more valuable in the sense that the marginal gain in reporting accuracy through auditing becomes greater. The higher marginal effect of \( a \) on the accuracy of reports makes auditing more attractive. For sufficiently low \( s \), this incentive outweighs the negative incentive of the lower error probability, and \( a^* \) overall increases. In economic terms this means that internal effort \( s \) not only has a negative effect on the marginal productivity of audit effort via the liability exposure, but has a positive effect on the productivity of the audit technology itself. This positive effect reflects the facilitation of audit procedures by (i) more sophisticated accounting and internal controls (e.g. investments of the firm into big-data computing), and (ii) the provision of organizational and testing assistance by the firm. Technically, the positive effect shows up as an increase in the slope \( \frac{\partial d(a,s)}{\partial a} \) in \( s \). Figure 4.2 illustrates the exponential audit technology for different levels of \( s \):

At the preceding stage of the game the firm chooses the level of internal effort \( s \), thereby

\(^{126}\) Proof: See appendix.

\(^{127}\) This intuition is particularly emphasized by Knechel / Willekens (2006), who state that ”the fact that audit fees are often higher cannot be explained within a production orientation since there is no obvious explanation why an auditor should bear more risk or perform more testing when the control environment is otherwise considered to be effective” (Knechel / Willekens (2006), p 1350).
implementing its demand for audit quality and overall financial reporting accuracy. The FOC for the firm’s effort choice reads

\[ \frac{dU^C(a^*, \alpha^*)}{ds} = \frac{k}{s^2} \frac{(I - V)}{V} - \frac{1}{c(s)} = 0. \]  \hfill (4.7)

The existence and uniqueness of an optimal choice \( s^* \) is stated in the following Remark 4.1:

Remark 4.1. For any increasing, non-concave cost-function \( c(s) \) there exists a unique equilibrium level of internal effort \( s^* \), which maximizes the firm’s expected utility \( U^C \).128

4.5 Comparative Statics

4.5.1 Internal Effort

Now assume that one of the exogenous model parameters increases at the margin. Since the auditor’s liability is the only parameter that can be directly controlled by a regulator,

128 Proof: See appendix.
let the damage payments $V$ increase by a small amount. For example, this increase could reflect the (perceived) tightening of auditors’ liability through the SOX. Note that the following analysis holds for changes in the other model parameters as well. In equilibrium, 

$$\frac{d^2 U^C(a^*, \alpha^*)}{dsdV} \equiv 0$$

must hold. Application of the implicit function theorem yields the following proposition:

**Proposition 4.1.** The equilibrium level of internal effort $s^*$ strictly decreases in the damage payments $V$.

Proposition 4.1 resembles a part of Proposition 1 in Pae / Yoo (2001). They state that a higher auditor liability discourages firm-side accounting activities, whereas it raises the equilibrium audit effort. However, there are two critical differences, that make a second look worthwhile. First, they formulate their model in a general manner and exogenously impose the regularity condition $\frac{\partial a^*}{\partial s \partial V} \leq 0$. Second, they exogenously impose strategic substitution between auditing and firm-side internal effort, whereas the two-faced exponential audit technology also exhibits strategic complementarity. In contrary, I start with a specific model and take any such conditions as by-products. This is important, because the nature of the auditor-client interaction is nothing given, but endogenously emerges from the relative resource investments. To see why and how these subtle changes matter, start by rewriting the client firm’s FOC (4.7) as

$$\frac{dU^C(a^*, \alpha^*)}{ds} = a^*e^{-(sa^*)^*}(1-p)(I-V) + s^*e^{-(sa^*)^*}(1-p)(I-V)\frac{\partial a^*}{\partial s} - c(s^*) \equiv 0. \quad (4.8)$$

129 Proof: See appendix.
130 See Proposition 1 in Pae / Yoo (2001).
131 The results of Pae / Yoo (2001) seem intuitive w.r.t. the effort-allocation, but are anything but general. A closer look reveals that they hinge on a series of other regularity conditions. Hence, there are plenty of good reasons to have a second look at this paper.
The first addend captures the direct marginal utility of a higher internal effort $s$, whereas the second addend captures the marginal utility of a change in the equilibrium audit effort $a^*$ due to a change in $s^*$. The latter captures the strategic effect of internal effort. The impact of a liability increase can be analyzed by differentiating the FOC w.r.t. $V$. Since the firm’s decision problem is strictly concave, the sign of the effect $\frac{ds^*}{dV}$ is given by the sign of

$$\frac{d^2 UC(a^*, \alpha^*)}{ds \partial V} = \frac{\partial^2 UC(a^*, \alpha^*)}{ds \partial V} + \left[ \frac{\partial^2 UC(a^*, \alpha^*)}{d^2 a} \right] \frac{\partial a^*}{\partial V} + \frac{\partial UC(a^*, \alpha^*)}{\partial a} \frac{\partial^2 a^*}{\partial s \partial V} \equiv \frac{d^2 UC(a^*, \alpha^*)}{ds \partial V}$$

$$\Leftrightarrow \frac{d^2 UC(a^*, \alpha^*)}{ds \partial V} = -e^{-(sa)^*} (1 - p)(a^* + s \frac{\partial a^*}{\partial s}) + e^{-(sa)^*} (1 - p)I \left[ 1 - (sa)^* - (s^*)^2 \frac{\partial a^*}{\partial s} \right] \frac{\partial a^*}{\partial V} +$$

$$\Leftrightarrow \frac{d^2 UC(a^*, \alpha^*)}{ds \partial V} = -s^* e^{-(sa)^*} (1 - p)I - k \frac{\partial^2 a^*}{\partial s \partial V}.$$

(4.9)

Three partial effects determine the direction of the overall effect $\frac{ds^*}{dV}$. Effect (I) is the direct effect of the damage payments $V$ on the firm’s FOC. This effect is negative, because the auditor does not observe the actual level $s^*$ before he is hired, whereas investors do observe $s^*$ before they set the cost of capital. Hence, any increase in $V$ lowers the cost-of-capital, while it leaves the fee unchanged. This makes accurate reporting relatively less important for the firm, and cuts down the incentive to invest in internal effort. However, in equilibrium the auditor’s conjecture $s^c$ is true, and thus the increase in the damage payments $V$, as well as any expected change in internal effort are priced, such that the
Note that if the fee is agreed upon after the auditor has observed $s^*$, then effect $(I)$ is null, because any direct incentive effect of $V$ via the cost of capital is exactly offset by a countervailing effect through the fee. In that scenario, $s$ represents activities and circumstances the auditor observes in the course of the client evaluation.

The second term in the first line of (4.9) depicts how the firm reacts to a direct change in the equilibrium audit effort. From Lemma 4.1 we know that the equilibrium audit effort $a^*$ directly increases in the damage payments. The analysis shows that in the present model the term in brackets is zero, which means that the firm does not care about the direct increase in audit effort. The reason for this is that in equilibrium the incentives of changes in the absolute value and the slope of the detection function induced by the direct effect $\frac{\partial a^*}{\partial V}$, exactly level each other out. More precisely, for currently sufficiently low (high) values of $s$, a marginal increase in $a^*$ increases (decreases) the marginal effectiveness of internal effort. However, the incentive to exert more (less) $s^*$ through the impact of a higher equilibrium audit effort on the slope coefficient $\frac{\partial U_C}{\partial s}$ is offset by the incentive to cut down (increase) internal effort because the equilibrium detection probability is currently sufficiently high (low). Formally, the necessary and sufficient condition for the term in brackets to be zero reads $\frac{\partial a^*}{\partial s} = -\frac{\frac{\partial^2 U_C}{\partial s \partial a}}{\frac{\partial^2 U_C}{\partial a^2}}$.

Generally, this condition holds for all auditor-client models in which the auditor’s and the client’s incentives to exert effort are aligned. This is certainly the case if the auditor is assumed to be the sole defendant due to client bankruptcy, or both the auditor and the firm are insured against damage claims. However, the condition does not hold in case of an effective damage apportionment under the following conditions:

132 In equilibrium, the increase in the fee due to the higher liability is exactly offset by a simultaneous decrease in the cost of capital $\alpha^*$. Thus the direct monetary effect is null, and only the incentive effect of liability remains. Laux / Newman (2010) call this mechanic “triangle effect” and illustrate it with a situation in which three persons each hand a ten dollar banknote to the person on their left; see Laux / Newman (2010), p 268.

133 It can be easily verified by $\frac{\partial^2 d(a^*)}{\partial d^2} = \frac{a^* e^{-a^*} - (a^*)^2 e^{-(a^*)}}{(a^*)^2 e^{-(a^*)}}$ and $\frac{\partial a^*}{\partial s} = \frac{1 - (s^*)^2}{(s^*)^2}$, that this condition holds in the present model. Pae / Yoo (2001) present a similar condition in their paper; see Pae / Yoo (2001), p 345.
proportionate liability, such as it is modeled in the first dissertation paper. Nevertheless, the results of this paper still go through, as long as the effort incentive provided by $\partial a^*/\partial V$ is not too strong.  

Now turn to the last addend in (4.9). In a model where the firm is generally interested in keeping the error probability low, the effect $\partial U/\partial a$ is positive, that is external auditing is a conventional complement to internal effort. Hence, it remains to determine the sign of $\partial^2 a^*/\partial s \partial V$. For the exponential audit technology assumed, this effect is strictly negative in equilibrium. The economic interpretation is as follows: Higher damage payments directly incite the auditor to work harder. Ceteris paribus, the higher audit effort weakens the cost-efficiency of the collaboration, or even causes testing redundancies. Thus the firm’s internal effort becomes less effective in triggering a high audit effort, respectively dampens the audit effort even stronger. In anticipation of this effect, the firm reduces its equilibrium effort $s^*$ at the first-stage.

The critical difference to Pae / Yoo (2001) is that this effect is a natural outcome of the multifaceted audit technology $d(a, s) = 1 - e^{-sa}$, and not an exogenously imposed regularity condition. To illustrate part (ii) of Proposition 4.2 replace the exponential detection function with the square root function $d_m(a, s) = \sqrt{sa}$ for the moment. This function looks very similar to the exponential function. It is concave and strictly increasing in both arguments, and therefore suitable to describe the audit technology. Figure 4.3 shows this function for different levels of $s$.

Then we have

$$\frac{dU^A}{da} = 0 \iff \frac{s(1-p)V}{2\sqrt{sa}} = k \iff a^*_m = \frac{s(1-p)^2V^2}{4k^2}. $$

---

134 See the proof of Remark 4.2 in the appendix.

135 Note that this function is not naturally bounded between zero and one, and thus requires a set of parameter restrictions. W.l.o.g. I assume that the parameter set is such that the equilibrium values of $s^*$ and $a^*$ in the modified model are between zero and one.
It can be easily verified that the condition \( \frac{\partial a^*}{\partial s} = -\frac{\partial^2 UC}{\partial s \partial a} \) still holds for this function.\(^{136}\) Since

\[
\frac{\partial^2 a^*_m}{\partial s} = \frac{2(1 - p)^2 V}{4k^2} > 0
\]

is clearly positive, the effect \( \frac{ds^*}{dV} \), as given by the first line of (4.9) is positive. This means that the auditor can utilize the firm’s effort more efficiently if his effort is higher. The liability increase is the kind of "kick" that enables this additional cost-efficiency. Because the firm is the first-mover of the game, this effect positively spills over to the firm and increases the marginal effectiveness of internal effort \( s \). As a result, higher damage payments enhance the firm’s governance system.

The following Proposition 4.2 generally summarizes how exogenous shocks affect the firm-side effort in the considered sub-class of sequential auditor-client games with complete information.

**Proposition 4.2.** In sequential auditor-client models with complete information, in which

1. \( \frac{\partial a^*}{\partial s} = -\frac{\partial^2 UC}{\partial s \partial a} \) holds, and
2. internal effort and external auditing are

\(^{136}\) The proof is omitted for brevity.
conventional complements ($\frac{\partial U_C}{\partial a} > 0$), an increase in the damage payments $V$

(i) decreases the equilibrium firm-side internal effort if $\frac{\partial^2 a^*}{\partial s \partial V}$ is negative, and

(ii) increases the equilibrium firm-side internal effort if $\frac{\partial^2 a^*}{\partial s \partial V}$ is sufficiently positive.

Proposition 4.2 may look a bit gnarled, but provides a general guideline how changes in the legal environment affect the equilibrium quality of the firm’s accounting, depending on the model specification. Condition (1) is realistic, if one assumes that both the auditor and the firm are equally interested in accurate financial reporting. Condition (2) restricts the proposition to models in which the auditor and the firm are both interested in a high accuracy of financial statements. Leaving everything else unchanged, in an earnings management or even fraud setting, the firm’s biasing activity and the auditor’s effort would be conventional substitutes ($\frac{\partial U_C}{\partial a} < 0$). Then the statement of Proposition 4.2 would be in the reverse way.

4.5.2 Audit Effort

Now turn to the equilibrium audit effort. Formally, the equilibrium reaction of the auditor reads

$$\frac{da^*}{dV} = \frac{\partial a^*}{\partial V} + \frac{\partial a^*}{\partial s} \frac{ds^*}{dV}.$$  \hfill (4.11)

The following Proposition 4.3 states how changes in the auditor’s liability affect the equilibrium level of audit effort.

Proposition 4.3. The equilibrium level of audit effort $a^*$

(i) increases in the damage payments $V$, if accounting and auditing effort are strategic substitutes, or weak strategic complements in equilibrium, and

\footnote{The proof directly follows from the first line of (4.9) and the above discussion of Proposition 4.1}
(ii) decreases in the damage payments $V$ if accounting and auditing effort are strong strategic complements in equilibrium.\textsuperscript{138}

Part (i) of the proposition is intuitive. In any case higher auditor liability directly encourages the auditor to exert more effort. Furthermore, we know that the client reduces $s^*$ if $V$ increases, which is why the auditor further increases his effort due to the strategic substitution $\frac{\partial a^*}{\partial s} < 0$. However, part (ii) is unexpected and goes against the finding of Pae / Yoo (2001). As figure 4.2 graphically illustrates (and the proof in the appendix shows), part (ii) is likely to occur, if the current level of audit effort is low and the equilibrium internal effort starkly declines in the liability payments. The economic intuition is as follows: If the damage payments are sufficiently low, the equilibrium audit effort is not only low, but also mostly driven by the strategic complementarity between internal effort and auditing. An increase in $V$ crops the benefit of strategic complementarity via the antecedent reduction of the firm’s accounting effort $s^*$. For a sufficiently strong current strategic complementarity, the resulting negative effort-incentive, depicted by the product in (4.11) outweighs the positive direct incentive $\frac{\partial a^*}{\partial V}$.

An implication of the above results is that increasing the threat of inaccurate reporting consequences can drive out cost-efficiency in auditing. An increase in legal liability makes the auditor care \textit{relatively} more about keeping the absolute error risk low than about exploiting synergies in the joint production of accurate financial statements. From the firm’s perspective, internal effort becomes less cost-efficient, which is why it cuts down its investment in own assurance activities. However, if this happens in an equilibrium in which the auditor \textit{absolutely} cares much about strategic complementarity, he actually reduces his effort. As a result, the equilibrium moves from a situation in which both players pull in the same direction, to a situation in which they both do their own thing. At a more general level, the reason for this unpleasant outcome can be found in dis-alignment

\textsuperscript{138} Proof: See appendix.
of the firm’s and the auditor’s incentive. While the firm cares about the auditor’s cost-efficiency, the auditor does not care about the firm’s cost-efficiency. Another interesting implication of Proposition 4.3 is that even if the two efforts are strategic complements, they might appear to be substitutes in equilibrium. This highlights the importance of distinguishing between an observable equilibrium association and the mostly unobservable strategic interaction. The result is also in line with empirical evidence of Felix / Gramling / Maletta (2001), who find that in the highly litigious U.S.-environment, where the audit effort is consequently expected to be high, the relation between external auditing and the internal auditing contribution is one of substitution.

4.5.3 Financial Reporting Accuracy

Putting the pieces together, we can now make the following statement about the aggregate impact of higher damage payments $V$ on the accuracy of financial statements:

**Corollary 4.1.** If internal effort and auditing effort are strong strategic complements in equilibrium, an increase in the damage payments $V$ decreases the accuracy of audited financial reports.$^{139}$

Corollary 4.1 adds to numerous findings highlighting that tightening auditor liability in order to increase the accuracy of published reports, might actually evoke the exact opposite outcome. Compared to prior findings of the auditor liability literature, the result is intriguing, because it proposes that the nature of the auditor-client collaboration can thwart effort-stimulating incentives. The analysis shows, that under reasonable conditions the two relevant characteristics of the auditor-client relationship are (i) the existence of a sufficiently strong effort complementarity between accounting and auditing ($\frac{\partial a^*}{\partial a} >> 0$), and (ii) the direction of the change in this complementarity ($\text{sgn}(\frac{\partial^2 a^*}{\partial a \partial V})$). To illustrate the

---

$^{139}$ The proof directly follows from Propositions 4.1 and 4.3.
point, turn back to the model variation with the square root audit technology \( d_m(a, s) = \sqrt{sa} \) from above.

Importantly, the square root audit technology exhibits strict strategic effort complementarity, that is

\[
\frac{\partial a^*_m}{\partial s} = \frac{(1 - p)^2 V^2}{4k^2} > 0
\]

for all levels of \( s^* \). Recall from above that \( \frac{\partial^2 a^*_m}{\partial s \partial V} > 0 \), that is the complementarity is enforced by a harsher auditor liability. *Ceteris paribus*, this means that the equilibrium accounting effort increases in \( V \left( \frac{\partial a^*_m}{\partial V} > 0 \right) \). Because the direct auditor incentive \( \frac{\partial a^*_m}{\partial V} = \frac{2s^*(1-p)^2V}{4k^2} > 0 \) is positive, the aggregate equilibrium change of audit effort, as given by (4.11) is clearly positive. It follows that the accuracy of audited financial reports strictly increases in the damage payments \( V \).

We can now summarize the findings in the following general proposition:

**Proposition 4.4.** In sequential auditor-client models with complete information, in which

(i) the condition \( \frac{\partial a^*_m}{\partial s} = \frac{\partial^2 a^*_m}{\partial s \partial a} \) holds, and (2) internal effort and external auditing are conventional complements (\( \frac{\partial U_C}{\partial a} > 0 \)), an increase in the damage payments \( V \)

(i) increases the equilibrium financial reporting accuracy if internal effort and audit effort are strategic complements (\( \frac{\partial a^*_m}{\partial s} > 0 \), and the complementarity is strongly reinforced (\( \frac{\partial^2 a^*_m}{\partial s \partial V} >> 0 \)).

(ii) increases the equilibrium financial reporting accuracy if internal effort and audit effort are strategic substitutes (\( \frac{\partial a^*_m}{\partial s} < 0 \), and the substitution is reinforced (\( \frac{\partial^2 a^*_m}{\partial s \partial V} < 0 \)).

Corollary 4.1 and Proposition 4.4 shed new light on the impact of exogenous shocks, such as a tightening of auditors’ liability, on financial reporting accuracy. It shows that

\[140\] The proof of the proposition directly follows from the first line of condition (4.9) and equation (4.11).
the intuitive statement of Pae / Yoo (2001), that internal effort and external auditing substitute each other in equilibrium, is not a “natural law” of sequential auditor-client interaction. Instead it follows from very specific model assumptions.

The following figures 4.4, 4.5 and 4.6 illustrate the above findings. All three figures depict the equilibrium internal effort \( s^* \), the equilibrium audit effort \( a^* \), the equilibrium detection probability \( d(a^*, s^*) \), and the equilibrium strategic effort-interaction \( \frac{da^*}{ds} \) as functions of the damage payments \( V \). The upper of the two bars indicates the regions in which a plain regression of internal effort on audit effort would yield a positive (negative) coefficient on the explanatory variable. The lower bar indicates the regions where internal effort in fact causes positive (negative) changes in audit effort. Figure 4.4 is based on the negative exponential detection function \( d(a, s) = 1 - e^{-sa} \). Figure 4.5 is based on the square root function \( d(a, s) = \sqrt{sa} \). Finally, figure 4.6 is based on the multiplicative detection function \( d(a, s) = 1 - (1 - s)(1 - a) \). This latter function is frequently used in the auditing literature, and likely the kind of function Pae / Yoo (2001) had in mind.\(^\text{141}\)

It is quite obvious that in general, strategic effort-substitution (complementarity), and an observable association are two completely different pairs of shoes.

4.5.4 Heterogeneous Stakeholder Interests

Empirical studies use either supply- or demand-side arguments to explain the overall ambiguous evidence. An advantage of the game-theoretic analysis is that it simultaneously analyzes supply-and demand-side effects, which (under the assumptions imposed on the bargaining power and the game sequence) form a market equilibrium. More precisely, the firm has a demand for a certain degree of financial reporting accuracy, because it

\(^{141}\) It should be noted that both the square root model and the multiplicative model are based on quadratic effort costs for both the auditor and the firm to make the respective optimization problems concave. The exponential model uses the parameter specification \( k = 2, I = 40, c = 2.5, p = 0.6 \). The square root model uses the parameter specification \( k = 1, I = 17, c = 2, p = 0.6 \). The multiplicative model uses the parameter specification \( k = 5, I = 30, c = 5, p = 0.8 \).
Figure 4.4: Equilibria for the exponential model specification

Figure 4.5: Equilibria for the square root model specification
internalizes the auditor’s and new investors’ interests. The firm then uses the level of
internal effort to make the auditor supply the desired audit effort. However, the base
model is narrow in the sense that all stakeholders have the same interest as the firm itself,
that is to minimize the probability of bad investment at the lowest cost. Both the internal
effort $s$, and audit effort $a$ exclusively serve this interest.

This section picks up the argument brought forward by Knechel / Willekens (2006) and
Hay / Knechel / Ling (2008). They argue that the diversity in a firm’s stakeholders’
interests elicits equilibrium complementarity, because investments in one governance in-
strument shift the costs and benefits of better corporate governance, and create positive
externalities on other instruments. As a consequence, a demand for complementary in-
vestments in other governance instruments arises to offset this shift, respectively to exploit
the externalities. To analyze whether such effects exist in a market equilibrium, hetero-
genous stakeholder interests have to be incorporated into the model. To this end, the
above base model is modified in the following way:

Assume that the firm consists of two players. First, there is the CEO who initially is the
sole owner of the firm. The CEO has the ultimate responsibility for the firm’s operations, investments and financial reporting and thus sets up the internal corporate governance system.\textsuperscript{142} Second, there is a non-owning manager who runs the firm’s ongoing operations. For simplicity, assume that the observable profit from the firm’s ongoing operations can only take the two values $x_H > 0$ and $x_L = 0$. The manager can either put in productive work, or slack off. If he decides to work, then an internal control system, which is part of the internal effort investment $s$ monitors the production process and timely informs the manager about unfavorable developments. For simplicity, assume that in case the manager works, the probability that the outcome is $x_H$ equals $s$.\textsuperscript{143} This means, that now $s$ plays a dual role in increasing the firm’s overall profitability. On the one hand, it prevents bad new investments, which is what external stakeholders, such as the new investors and the auditor, care about. On the other hand, it increases the correlation between productive effort and operational profit, which is what the operational manager cares about. This modeling captures the fact that investments in corporate governance, such as better internal controls, also serve internal purposes, whereas the audit is addressed to external stakeholders only.

If the manager decides to slack off, the chances are high that the outcome is $x_L$. Only with a sufficiently small probability $q < s$ the manager is lucky and the outcome is $x_H$. Because the manager’s effort level itself is unobservable, and working hard creates private costs of $v_H$ to him, the owner-manager has to provide work-incentives through the compensation contract. The contract is written after the owner-manager has set up the internal control system through his investment $s$, and specifies the payments to the manager contingent on the observed outcome. Let these payments be denoted by $g_H$ and $g_L$. For simplicity, let the manager be protected by limited liability $g_L \geq 0$. Figure 4.7 sketches the sequence of the modified model.

\textsuperscript{142} The CEO does the same as ”the firm” in the base model.

\textsuperscript{143} If one narrowly interprets the whole $s$ as investment into the internal control system, $s$ determines the probability that a diligent manager detects diversion by employees.
In summary, the difference between the base model and the modified model is that the owner-manager’s demand for financial reporting accuracy (an \((s^*, a^*)\)-tuple) is driven by heterogeneous stakeholder interests. Since the new investors’ choice of \(\alpha(s, a)\) and the auditor’s choice of \(a^*(s)\) is decoupled from the agency problem between the owner-manager and the manager, the optimal contract is given by the solution to the following program:

\[
\begin{align*}
\max_{g_H, g_L} & \quad p(1 - \alpha(s, a))Z - W - cs + sx_H - (sg_H + (1 - s)g_L) \\
\text{s.t.} & \quad g_Hs - v_H + (1 - s)g_L \geq 0 \tag{4.13} \\
\text{s.t.} & \quad g_Hs - v_H + (1 - s)g_L \geq qg_H + (1 - q)g_L \tag{4.14} \\
\text{s.t.} & \quad g_L \geq 0 \tag{4.15}
\end{align*}
\]

W.l.o.g. assume that the parameters are such that the owner-manager wants to hire a manager and wants him to work hard instead of slacking off. Because the owner does not benefit from paying the manager more than necessary to make him work hard, the second constraint \(\text{(4.14)}\) is binding, that is
Inserting this expression into the first constraint (4.13) and solving for $g_L$, one can easily see that $g_L$ would be negative. This means, the third constraint (4.15) is binding and $g_L = 0$, whereas the first constraint (4.13) is not binding. It follows that the compensation for the high outcome is equal to

$$g_H = \frac{v_H}{s - q}.$$ (4.16)

At the first stage of the game the owner-manager chooses the level of internal effort, now also taking into account its implications for the contracting with the manager. Formally, the owner-manager chooses the effort level $s$ that maximizes

$$U^{C}_{mod} = p(1 - \alpha(s,a))Z - W - cs + s(x_H - \frac{v_H}{s - q}).$$ (4.17)

Under a parameter set that ensures strict concavity of (4.17), the equilibrium level of internal effort is given by the unique solution of the FOC

$$\Leftrightarrow \text{FOC} + \left[ \frac{x_H - \frac{v_H}{s - q} + \frac{sv_H}{(s - q)^2}}{s - q} \right] = 0,$$ (4.18)

whereas the component $\text{FOC}$ is the FOC of the base model, as given by (4.7). Because the term in brackets is strictly positive, the equilibrium level of internal effort in the modified model is strictly greater than in the base model. The following lemma summarizes how the equilibrium level of internal effort depends on the severity of the agency problem.
Lemma 4.2. The equilibrium internal effort $s^*$ strictly increases in the manager’s effort costs $v_H$, and the probability $q$.\footnote{The proof directly follows from condition (4.18)}

As intuition suggests, the more severe the agency conflict between the owner and the manager is, the higher the internal effort investment will be, since this allows the owner to cut down the incentive-compatible compensation. The magnitude of the bracket term captures how much the interests $s$ has to serve diverge from the pure loss-prevention interest in the base model.

Now return to the comparative static analysis. Because the bracket term in (4.18) does neither directly, nor indirectly depend on $V$, the equilibrium $s^*$ still strictly decreases in the liability payments $V$ (see Proposition 4.1), whereas the term $\frac{da^*}{dV}$ remains identical to the base model. However, because the modification affects the equilibrium level of $s^*$, the agency problem spills over to the auditor’s second-stage decision about audit effort. This leads to the following two propositions:

**Proposition 4.5.** The more severe the agency problem is ceteris paribus (and consequently, the less important the investment problem is relatively),

(i) the more likely internal effort and external auditing are strategic substitutes in equilibrium ($\frac{ds^*}{ds} < 0$), and

(ii) the more likely an increase in the liability payments $V$ increases the equilibrium audit effort $a^*$.\footnote{Proof: From Lemma 4.2 we know that $s^*$ is strictly increasing in both $v_H$ and $q$. Part (i) then directly follows from Lemma 4.1. Part (ii) immediately follows from (4.28)}

**Proposition 4.6.** An increase in the severity of the agency problem strictly increases both the internal effort level $s^*$ and external audit effort $a^*$, and therefore the accuracy of financial reports, whenever the two tasks are strategic complements in equilibrium ($\frac{ds^*}{ds} > 0$).\footnote{Proof: From Lemma 4.2 we know that $s^*$ is strictly increasing in both $v_H$ and $q$. The result then directly follows from Lemma 4.1}
The intuition of the propositions is as follows: An increase in the severity of the agency problem between the owner-manager and the non-owning manager relatively drives out the demand for cost-efficient accurate financial reporting. Therefore the owner-manager increases $s$ to reinforce the work-incentives for the manager, accepting that the auditor will cut down his effort due to the low probability of liability. As the demand-forces that drive audit effort and internal effort digress when the agency problem becomes more severe, the owner-manager cares relatively less about the cost-efficient collaboration with the auditor when there are changes in the auditor’s legal liability incentives. Consequently both players start to do their own thing, and not only is there unobservable strategic substitution, but also an observable divergence of the equilibrium resource investments. In summary, Proposition 4.5 states that a divergence in the interests the various control instruments have to serve tends to eliminate the joint incentives of legal liability on the auditor and the firm, because it turns down the cost-efficient sharing of work.

Proposition 4.6 however makes a different statement. It claims that the increase in the agency problem itself, respectively the divergence of interests itself, causes higher audit quality, whenever internal effort and external auditing are strategic complements. An increase in the agency problem raises the demand for internal effort, because the manager can then be incentivized more efficiently. Under strategic complementarity, this simultaneously raises the demand for audit effort, because the firm can gain additional financial reporting accuracy, and thereby serve the loss-prevention interest more effectively if it puts more resources into the auditing process. In that case the financial reporting accuracy unambiguously increases. However, if internal effort and the audit are strategic substitutes, then the opposite is true and the equilibrium audit effort declines as the agency problem within the firm becomes more severe.

Proposition 4.6 is perfectly in line with the empirical findings and arguments of Knechel/Willekens (2006) and Hay/Knechel/Ling (2008). They document a complementary relation between the institutions in unregulated corporate governance environments. In
my model such an unregulated environment would be represented by low control costs $c$. In reality these are likely to occur when agency problems within the firm are innocuous (low $v_H$ or low $q$), or the economic environment is prosperous (high $p$). We know from Lemma 4.1 that under these conditions strategic complementarity is likely to prevail in equilibrium. Knechel / Willekens (2006) also find that in environments where the internal control investments are high due to regulation, the positive demand externality on the external audit is driven out, and the relation becomes one of substitution. This is exactly what Proposition 4.6 predicts.

The model extension shows, how various (demand-side) arguments can be integrated into an equilibrium production-view model. The existence of the agency problem augments the scope of goals the two efforts have to serve, from a purely external interest to both external and internal interests. Now, the investment in an $(s,a)$-tuple the firm effectively makes, not only aims at keeping the accuracy of financial reports high, but also increases the manager’s productivity. Hence, the effective production-technology is no longer $d(a, s) = 1 - e^{-sa}$, but a composition of $d(a, s)$ and $sx_H$. Since $sx_H$ is linear in $s$, and only regards the firm, the aggregate production function still exhibits the distinctive double-edged strategic effect $\frac{da^*}{ds}$. In the same manner, one could implement the argument of Goodwin-Stewart / Kent (2006), that the investment into internal governance is part of a commitment to a strong overall corporate governance, and that the shareholders then also demand high quality external audits. This would be incorporated into the model by adding a payoff component $z(a, s)$ that captures the benefits of having a strong overall corporate governance. The aggregate production technology would then be a composition $d(a, s) \circ z(a, s)$. The shape of the production function would then presumably be neither exponential, nor a square root or multiplicative, but more complex. However, regardless of the actual shape of the function, if the interaction between the auditor and the client firm is sequential, the resource allocation will boil down to the two equilibrium characteristics $\frac{da^*}{ds}$ and $\frac{d^2a^*}{dsds^*}$. 
4.6 Empirical Application

In this section I subject the proposed theory to empirical scrutiny. In essence, the analysis predicts that when corporate governance and internal controls are heavily regulated, and the threat of litigation is sufficiently high, internal effort and external auditing are substitutes. However, for rather weak internal corporate governance and a sufficiently mild liability environment, the model predicts a complementary relation. I test whether these predictions hold in reality, using two firm panels from Germany and the U.S. It is well-known that the litigation environment in the U.S. is considerable harsh, whereas it is very mild in the German-speaking countries. This is most prominently documented by La Porta / Lopez-de Silanes / Shleifer (2006), who report a liability index of 1.00 (which is the maximum) for the U.S., whereas they report a liability index of 0 (the minimum) for Germany. Moreover, the notorious Sections 302 and 404 of the SOX dramatically raised the internal control compliance and documentation requirements for firms, threatening severe penalties to non-complying firms. Numerous companies publicly groaned under the additional costs of complying with the new regulations. Shortly after SOX, the European Union enacted the 8th Company Law Directive, which is fondly called "EuroSOX". Although the 8th Directive basically pursues the same goal as the SOX, namely to restore investors’ confidence in corporate disclosures, it is considerably milder in various aspects. Altogether, the weaker governance regulation and the less rigorous liability regime lead to the presumption that German firms spend less resources on corporate governance and internal control compliance than U.S. firms, whereas auditors are by far less exposed to liability. These two environments are therefore representative of...

147 I am indebted to Wayne Landsman for encouraging me to do an empirical analysis, and providing me with the necessary toolkit during my research visit at UNC Chapel Hill in the fall semester 2017.
148 Also the other indices of securities markets regulation reported by La Porta / Lopez-de Silanes / Shleifer (2006) are on the top end for the U.S., but on the bottom end for Germany; see La Porta / Lopez-de Silanes / Shleifer (2006), pp 15 et seq.
149 For example, under SOX it is "comply or die", whereas under the 8th Directive it is "comply or explain".
the two interesting institutional settings the theoretical analysis refers to.

4.6.1 Research Design

Sample Selection

I use two different firm panels for the years 2005 to 2014. The basis for the U.S.-sample are the firms listed in the S&P 500 index within this period. Starting out with 5,050 firm-year observations from 505 different firms, I delete observations which either lack some of the regression variables, or regard firms in the financial industry. I also delete firms with audit fees over a hundred million U.S.-Dollars, since there are some comma faults in the raw data. Correspondingly, the German sample comprises the firms listed in the Prime All Share index between 2005 and 2014. Starting with 3,230 firm-year observations from 323 different firms, I apply the same elimination procedure as for the U.S.-sample. In addition, I eliminate observations of firms which do not report their financial data in Euro. These eliminations yield a final U.S.-sample of 2,877 observations coming from 385 different firms, and a final German sample of 1,167 observations coming from 207 different firms.

Model Specification

The hypothesis to be tested regards the sign of the coefficient on a proxy variable for the internal effort $s$ in regressions of a proxy for audit quality on the internal effort proxy. Based on the above analysis, I expect a positive sign for the German sample and a negative sign for the U.S.-sample. Since the firm’s decision about the level of internal effort $s$ is endogenous, a conventional OLS regression yields biased and inconsistent coefficient estimates. To overcome this problem, I use an instrumental variable 2SLS approach.

\footnote{Some firms would have paid billions for the audit according to the raw data. In 2014 the maximum audit fee of the Fortune 100 companies which are not in the financial industry amounted to 91 million U.S.-Dollars (General Electric Corp.).}

\footnote{A Durbin-Wu-Hausman test was run to document the endogeneity problem (untabulated).}
In essence, I follow the approach of Hay / Knechel / Ling (2008). However, Hay / Knechel / Ling (2008) stay within the New Zealandian environment, and compare firms before and after major institutional changes (1995 and 2005), whereas I compare firms from two different institutional environments over an eleven-year time-window.

All variables are obtained from the Thomson Reuters Worldscope, I/B/E/S and ASSET4 ESG databases through Datastream. The endogenous variable to be explained is audit quality, which is neither directly observable, nor quantifiable. Following prior literature, I use the natural logarithm of the total fees paid to the auditor (in thousands) \( \text{LNFEE} \) as proxy for the dependent variable audit quality (Datastream item WC01801). The primary explanatory variable of interest is the firm-side effort \( s \).

Before racking our brains about proxy variables for \( s \), we need to establish a link between the equilibrium audit quality \( a^* \) and audit fees. Consistent with practice (London Economics/ Ewert (2006)) and prior theoretical literature, I assume that the audit market is perfectly price-competitive, that is auditors receive zero rents in equilibrium. Setting the auditor’s reservation utility equal to zero gives the equilibrium audit fee

\[
W(s^*) = \frac{k}{s^*} \left[ 1 + \ln \left( \frac{s^*(1-p)V}{k} \right) \right],
\]

which can be shown to be strictly decreasing in \( s \). This means that for equilibrium strategic complementarity of efforts, the audit fee \( W(s^*) \) does not follow the audit quality \( a^* \). Under strategic effort complementarity, the auditor benefits so much from the increase in the marginal effectiveness of his technology, that the decrease in liability risk outweighs the increased effort costs (this is why he increases his effort). However, in what form the client firms enjoy the gain in cost-efficiency is a different question. In the base model,

\[^{152}\text{In their survey London Economics/ Ewert (2006) document that a large majority of auditors is not even able to fully price the litigation risk into the fees.}\]

\[^{153}\text{See the proof of Remark 4.2 in the appendix.}\]
hiring the auditor is a necessary duty to obtain financing. The firm’s (and the auditor’s) only goal is to get an accurate, unqualified opinion at the lowest cost possible. However, in reality the statutory audit is only a part of the services firms purchase from audit firms. A substantial amount of the fees typically accrues to services that are not directly related to the actual statutory audit, such as consultancy or tax-related services. If the auditor spends more time and resources on the statutory audit, he gains deeper insights into the firm’s business and accounting practices, which make the non-audit services (NAS) more effective (“knowledge spillovers”). I capture this synergy effect by modifying the client firm's and the auditor’s expected utility functions in the following simple way:

\[ U_{C_{\text{mod}}} = p(1 - \alpha)Z - W(s^c) - cs + ra - F(r), \]  

(4.19)

and

\[ U_{A} = W(s^c) - (1 - p)(1 - d(a,s))V - ka + F(r) - \frac{r^2}{2}. \]  

(4.20)

After spending effort \( s \) on the internal governance, the firm purchases NAS \( r \) from the auditor.\(^{154}\) The NAS increase the final payoff to the client-firm at a rate that increases in the amount of audit effort exerted by the auditor. For example, \( ra \) captures increases in the operative profitability, or reductions in tax payments through external consulting. Providing NAS causes costs of \( \frac{r^2}{2} \) to the auditor which the fee \( F(r) \) exactly covers. Solving the modified model yields the following results:

**Remark 4.2.** In the model with NAS, an increase in the liability payments \( V \)

\( (i) \) increases the equilibrium internal effort \( s^* \), but decreases the equilibrium audit quality \( a^* \) under strategic effort-substitution \( (\frac{\partial a^*}{\partial s} < 0) \),

\(^{154}\) Technically, it does not matter whether the firm purchases the NAS simultaneously to, or after exerting the internal effort \( s \).
(ii) increases both the equilibrium internal effort $s^*$, and the equilibrium audit quality $a^*$ under sufficiently strong strategic effort complementarity ($\frac{\partial a^*}{\partial s} > 0$) if the initial investment amount $I$ is sufficiently large, but the marginal direct audit costs $k$ are sufficiently small.

(iii) Under strategic effort-substitution ($\frac{\partial a^*}{\partial s} < 0$), the total fees paid to the auditor decrease in internal effort $s$.

(iv) Under sufficiently strong strategic effort complementarity ($\frac{\partial a^*}{\partial s} >> 0$), the total fees paid to the auditor increase in internal effort $s$ if the initial investment amount $I$ is sufficiently large, but the marginal direct audit costs $k$ are sufficiently small.$^{155}$

Remark 4.2 is the theoretical basis for the empirical model. As the base model, it predicts a negative association between audit quality and internal governance for sufficiently high $s^*$, but a positive association for sufficiently strong $s^*$. The additional conditions of sufficiently small marginal direct audit costs, but sufficiently high investment $I$ for the latter case are rather mild, since all firms in both samples are very large, and audited by Big-N firms. For the case of strategic effort-substitution, the equilibrium total audit fees decrease in $s^*$, since the auditor cuts down both the audit effort $a^*$ and the NAS. However, for sufficiently small $s^*$ and thus sufficiently strong strategic effort complementarity, audit fees increase in $s^*$, since the increase in audit effort boosts the firm’s demand for NAS. Then the increase in the non-audit fees outweighs the reduction of the audit-related part of the fee.

The proxy variable for $s$ is AUDSCOR, which is the unweighted average of four different corporate governance percentage scores available from the ASSET4 ESG database. COMEXSCOR (Datastream item CGBFO03S) is the score for the expertise of the firm’s audit committee. It is higher if the audit committee has more members and at least one financial expert in the sense of the SOX. COMINDSCOR (Datastream item CGBFO01S)

$^{155}$ Proof: See appendix.
is the score for the independence of the audit committee from the firm. It is higher if there is a higher percentage of independent board members on the audit committee. ACCONTRSCOR (Datastream item ECSLO17S) is a score that refers to accounting controversies of the firm. It is higher, if there is frequent media coverage about aggressive and non-transparent accounting of the firm. Finally, RESTSCOR (Datastream item ECSLO10S) is a score that reflects whether the firm is in the process of a material earnings restatement. **AUDSCOR** ranges between zero and a hundred, whereas higher values represent better internal governance.

The problem with AUDSCOR is that in the German sample for 791 of the firm-year observations (67.78%) the corporate governance scores used to calculate AUDSCOR are missing. In the U.S.-sample the problem is less severe, with only 180 (6.26%) score-bundles missing. The natural solution would be to eliminate all firm-year observations with missing corporate governance scores (complete case analysis). However, eliminating data points (i) reduces statistical power and, more importantly (ii) biases the coefficient estimates if the missing data is not completely randomly distributed. A logit regression of an indicator for the missing data points on the other regressors (untabulated) shows that the variables are significantly (1%-level) more often missing for (i) earlier years in the sample, (ii) smaller firms, and (iii) firms with less analyst following. Importantly, the corporate governance information reflected in the ASSET 4 ESG scores is not information the firms themselves decides to disclose or not to disclose based on the actual values of the scores. It is information the database provider either retrieves from the firms’ reporting (which is rather extensive, since the indices comprise the largest firms of the respective economies), or from other sources (such as information on accounting controversies). Therefore the data is said to be missing at random (MAR), that is the missingness is related to some of the other regression variables, but unrelated to the un-

---

**Footnote:** Since ACCONTRSCOR and RESTSCOR have negative polarity, I use their inverse to get a positive polarity for the aggregate proxy AUDSCOR.
observed value of the missing data point itself. The standard technique of dealing with MAR-data is multiple imputation, originally introduced by Rubin (1987). In essence, multiple imputation derives a conditional distribution of AUDSCOR given the set of other variables from the complete observations. Then, \( m \) values for each of the missing data points are drawn from the distribution which gives \( m \) complete datasets. These datasets are separately analyzed with the regression model of choice, and then pooled for making statistical inferences. Because conceptually multiple imputation yields unbiased and more efficient estimators under the MAR assumption, it is considered to be superior to a complete case analysis. Specifically, I use multiple imputation by chained equations (MICE) with \( m = 10 \).

Following prior literature, I pick the following control variables:

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Measure</th>
<th>Expected Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>LN_ASSETS</td>
<td>Natural logarithm of total assets (in thousands)</td>
<td>+</td>
</tr>
<tr>
<td>OPINION</td>
<td>1 if unqualified audit opinion; 0 otherwise</td>
<td>-</td>
</tr>
<tr>
<td>INVREC</td>
<td>(Inventory + Receivables)/Total assets</td>
<td>+</td>
</tr>
<tr>
<td>LNEMP</td>
<td>Natural logarithm of the number of employees</td>
<td>+</td>
</tr>
<tr>
<td>FOR</td>
<td>Foreign assets/ Total assets</td>
<td>+</td>
</tr>
<tr>
<td>CURRAT</td>
<td>Current ratio</td>
<td>-</td>
</tr>
</tbody>
</table>

See Rubin (1976) for the generally accepted classification of missing data. Rubin (1987) showed that for datasets with MAR-data, multiple imputation gives unbiased estimators that properly account for the uncertainty due to the fact that the true values are missing. Multiple imputation is very popular in medical statistics. Conceptually, it is similar to the nonparametric bootstrap, which is also used in accounting research when data is missing or distributional properties are unknown. However, the bootstrap is a more simple and direct procedure, as data is generated by repeated draws from the distribution of the observed data points. Multiple imputation is more advanced in that the imputed values are drawn from a conditional distribution of the variable with missing values, which is derived given the distribution of observed data points. For a comparison of the two techniques see Efron (1994), pp 470 et seq.

See for example Newgard (2006) or Sterne et al. (2009). MICE is the state-of-the-art multiple imputation procedure. Rubin (1987) shows that an estimate derived from \( m \) imputations is approximately \( \frac{1}{1 + \frac{F}{m}} \) as efficient as the estimate derived from infinitely many imputations, whereas \( F \) denotes the fraction of missing data. For the German (U.S.) sample \( m = 10 \) imputations result in a relative efficiency of 93.65% (99.38%).
It is well known that firm size as measured by LNASSETS drives audit fees, since larger firms require more audit effort due to the greater number of transactions, and the greater variety of risks. A qualified opinion indicates that the auditor uncovered major flaws in the financial statements, or encountered severe impediments during the audit process. Both kinds of events result in significantly more audit hours, and thus higher fees. Inventory and receivables are financial statement positions that are more expensive and riskier to audit, because they exhibit considerable margins of discretion. Therefore, I predict a positive relation between INVREC and LNFEES. LNEMP and FOR both capture the complexity of the audit process. LNEMP regards organizational complexity, whereas FOR rather regards financial statement complexity. Both signs are predicted to be positive. To some extent, the variables INVREC, LNEMP and FOR reflect the audit cost parameter \( k \).

Finally, CURRAT captures the financial risk of the firm, whereas a higher current ratio value indicates a lower risk, and thus should be associated with lower audit fees. CURRAT shall capture the probability \( p \) of the model.

I use four instrumental variables to obtain estimates for the primary variable of interest AUDSCOR at the first stage of the regression model. ACOV is the mean of the total number of earnings-per-share estimates over the firm year, and reflects the firm’s analyst coverage. A high analyst coverage is primarily important for the firm because it reduces information asymmetries between the firm and its investors, thereby increasing liquidity ([Alfond / Berger (1999), Chang / Dasgupta / Hilary (2006)]), credibility of earnings ([Francis / Schipper / Vincent (2002), Frankel / Kothari / Weber (2006)]), and competitiveness ([Arya / Mittendorf (2007)]). In order to benefit from these effects, respectively to avoid adverse effects, the firm has to ensure a high accuracy of financial reports by maintaining strong governance mechanisms. The second instrumental variable is FREE. It reflects the percentage of the firm’s free float shares. All holdings below 5% are counted as free-float shares. Ownership concentration is very likely to have an effect on internal governance. However, the evidence on the direction is ambiguous. On the one hand,
large shareholders should be interested in protecting their substantial investments, and thus engage in active monitoring, respectively demand the installation of effective internal controls (Shleifer / Vishny (1986), Dechow / Sloan / Sweeney (1996)). On the other hand, ownership concentration might reduce firm-side investments in accurate reporting, as large shareholders have access to other information channels than financial reporting, and might be interested in using this information advantage against minor shareholders (Shleifer / Vishny (1997)). The third instrumental variable is LEVERAGE, which reflects a firm’s ratio of total debt to total assets. It is reasonable to assume that firms with higher leverage invest more into internal controls, and are more eager to provide accurate financial statements in order to get favorable financing conditions. In their survey of the empirical literature on the determinants of audit fees Hay / Knechel / Wong (2006) document that although frequently used in studies, the association between audit fees and firm leverage is often insignificant. They argue that this is because leverage is more directly related to internal governance than to the external audit, and thus use it as an instrumental variable. Fourth and finally, UTILITY is an indicator variable taking the value of one if a firm is in the utility sector (Worldscope general industry classification, item WC06010). Due to their importance to the general public, firms in this sector are typically subjected to dedicated regulatory authorities. The presence of these regulators is expected to improve corporate governance and increase internal control activity.\footnote{I do not include an indicator variable for the auditor size (Big N vs. Non-Big N) because all firms in the final sample were audited by the six biggest audit firms, with the Big-Four firms making up over 95\%.\footnote{For some French firms, Mazars is listed as the firm’s primary auditor. However, in France joint audits are mandatory since 1996, which makes Mazars the fourth biggest audit firm.}} The above ingredients form the following 2SLS regression model.

\footnote{See Hay / Knechel / Ling (2008).}
\[ \text{AUDSCOR}_{i,t} = \alpha_{i,t} + \beta_1 \times \text{ACOV}_{i,t} + \beta_2 \times \text{FREE}_{i,t} + \beta_3 \times \text{LEVERAGE}_{i,t} + \beta_4 \times \text{UTILITY}_{i,t} + \beta_5 \times \text{LNASSETS}_{i,t} + \beta_6 \times \text{OPINION}_{i,t} + \beta_7 \times \text{INVREC}_{i,t} + \beta_8 \times \text{LNEMP}_{i,t} + \beta_9 \times \text{FOR}_{i,t} + \beta_{10} \times \text{CURRAT}_{i,t} + \epsilon_{i,t} \] 

(4.21)

\[ \text{LNFEE}_{i,t} = \delta_{i,t} + \gamma_1 \text{AUDSCOR}_{i,t} + \gamma_2 \times \text{LNASSETS}_{i,t} + \gamma_3 \times \text{OPINION}_{i,t} + \gamma_4 \times \text{INVREC}_{i,t} + \gamma_5 \times \text{LNEMP}_{i,t} + \gamma_6 \times \text{FOR}_{i,t} + \gamma_7 \times \text{CURRAT}_{i,t} + \nu_{i,t} \] 

(4.22)

The coefficient of interest is \( \gamma_1 \). Based on the theoretical analysis, it is predicted to be positive for the German sample, but negative for the U.S.-sample.

### 4.6.2 Results

**Descriptive Results**

Tables 1a and 1b present the descriptive statistics for the two samples. Consistent with expectations, the internal governance variable AUDSCOR is on average around twenty points higher in the U.S., than it is within the German sample\(^{163}\). Unsurprisingly, German firms pay on average less for the audit, are smaller, and less complex than U.S. firms. They also have less analyst following, a lower percentage of free-float, and a lower leverage. However, they have a higher percentage of foreign assets, more current assets, and a higher current ratio. In terms of the model, this suggests that the RHS of condition \((4.24)\) is substantially higher for the U.S.-sample than for the German sample, whereas

\(^{163}\)The difference is statistically significant at the 1%-level. Note that the descriptive statistics for AUDSCOR only refer to the 2,697, respectively 376 complete cases.
for the LHS the opposite is true. This means that strategic complementarity is likely to prevail in the German context, which is a necessary condition to support the hypothesis.

Table 1a: Descriptive Statistics: S&P 500

<table>
<thead>
<tr>
<th>Panel A: Continuous Variables</th>
<th>Mean</th>
<th>Median</th>
<th>St. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNFEF</td>
<td>8.690</td>
<td>8.675</td>
<td>0.950</td>
<td>6.252</td>
<td>11.447</td>
</tr>
<tr>
<td>AUDSCOR</td>
<td>59.536</td>
<td>60.07</td>
<td>4.314</td>
<td>33.073</td>
<td>89.66</td>
</tr>
<tr>
<td>INVREC</td>
<td>0.210</td>
<td>0.188</td>
<td>0.144</td>
<td>0</td>
<td>0.799</td>
</tr>
<tr>
<td>FOR</td>
<td>12.720</td>
<td>4.94</td>
<td>18.436</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>CURRAT</td>
<td>1.884</td>
<td>1.53</td>
<td>1.269</td>
<td>0.14</td>
<td>12.3</td>
</tr>
<tr>
<td>ACOV</td>
<td>17.055</td>
<td>16</td>
<td>7.554</td>
<td>1</td>
<td>54</td>
</tr>
<tr>
<td>FREE</td>
<td>88.839</td>
<td>92</td>
<td>12.128</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>LEVERAGE</td>
<td>25.015</td>
<td>23.85</td>
<td>16.323</td>
<td>0</td>
<td>110.6</td>
</tr>
</tbody>
</table>

N = 2,877 (AUDSCOR: 2,697)

Panel B: Binary Variables

<table>
<thead>
<tr>
<th>YES %</th>
<th>NO %</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPINION</td>
<td>99.86%</td>
</tr>
<tr>
<td>UTILITY</td>
<td>10.57%</td>
</tr>
</tbody>
</table>

Table 1b: Descriptive Statistics: PRIME ALL SHARE

<table>
<thead>
<tr>
<th>Panel A: Continuous Variables</th>
<th>Mean</th>
<th>Median</th>
<th>St. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNFEF</td>
<td>6.820</td>
<td>6.503</td>
<td>1.601</td>
<td>3.761</td>
<td>11.208</td>
</tr>
<tr>
<td>AUDSCOR</td>
<td>39.552</td>
<td>42.406</td>
<td>11.617</td>
<td>23.725</td>
<td>63.108</td>
</tr>
<tr>
<td>LNASSETS</td>
<td>13.916</td>
<td>13.651</td>
<td>2.313</td>
<td>8.268</td>
<td>19.66</td>
</tr>
<tr>
<td>INVREC</td>
<td>0.334</td>
<td>0.336</td>
<td>0.161</td>
<td>0</td>
<td>0.865</td>
</tr>
<tr>
<td>LNEMP</td>
<td>8.410</td>
<td>8.426</td>
<td>2.191</td>
<td>2.565</td>
<td>13.292</td>
</tr>
<tr>
<td>FOR</td>
<td>27.331</td>
<td>22.030</td>
<td>23.959</td>
<td>0</td>
<td>164.16</td>
</tr>
<tr>
<td>CURRAT</td>
<td>1.935</td>
<td>1.580</td>
<td>1.421</td>
<td>0.21</td>
<td>16.42</td>
</tr>
<tr>
<td>ACOV</td>
<td>12.267</td>
<td>9</td>
<td>10.296</td>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>FREE</td>
<td>67.817</td>
<td>72</td>
<td>25.586</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>LEVERAGE</td>
<td>20.718</td>
<td>19.340</td>
<td>15.279</td>
<td>0</td>
<td>128.45</td>
</tr>
</tbody>
</table>

N = 1.167 (AUDSCOR: 376)

Panel B: Binary Variables

<table>
<thead>
<tr>
<th>YES %</th>
<th>NO %</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPINION</td>
<td>99.66%</td>
</tr>
<tr>
<td>UTILITY</td>
<td>4.20%</td>
</tr>
</tbody>
</table>

Tables 2a and 2b report the univariate Pearson correlation coefficients. Most of the coefficients are highly significant, but low absolute values indicate the absence of multicollinearity problems. The only exceptions are the two size-variables LNASSETS and LNEMP, which are positively correlated in both samples (S&P500: $\rho_{\text{LNASSETS, LNEMP}} = 0.571$;
Prime All Share: \( \rho_{(\text{LNASSETS, LNEMP}) = 0.926} \), and ACOV, which is strongly positively correlated with both LNASSETS and LNEMP in the German sample \( \rho_{(ACOV, LNASSETS)} = 0.794, \rho_{(ACOV, LNEMP)} = 0.731 \). However, separate regressions in which either LNEMP or ACOV were omitted gave only negligible changes in the regression coefficients.

Regression Results and Discussion

Tables 3a and 3b report the main results of the 2SLS regressions. Column (1) reports the results of the base model as specified above. Column (2) reports the regression with year-fixed effects included. Since the inclusion of time-fixed-effects is standard in panel regressions with longer time-windows, the results of model (2) are stronger. Standard errors are clustered at the firm-level.

In both model specifications, the coefficients on AUDSCOR are significantly negative for the U.S.-sample (10%--, respectively 5%-significance level), whereas they are significantly positive (5%-significance level) for the German sample. This means that taking into account the endogenous strategic relation between internal governance activities and external auditing, the observable association is substitution in the heavily regulated U.S.-environment, but complementarity in the laxer German environment. These findings support the theory outlined above. In the U.S.-sample all other coefficients except CURRAT have the expected signs and are mostly highly significant. Remember that CURRAT is used as a proxy for a firm’s business risk \((1 - p)\), which according to the model should negatively affect audit fees. However, the coefficient on CURRAT is positive (although not significant at the 10% level) in both specifications. One explanation is that the current ratio is driven upwards by short-term receivables, which is a balance sheet position auditors typically pay particular attention to. This is likely to (partially) offset the lower audit risk of firms with a high current ratio. In the German sample the coefficient on INVREC is negative (although also not significant at the 10%-level), whereas it should
<table>
<thead>
<tr>
<th>Str</th>
<th>Utilities</th>
<th>Option</th>
<th>Leverage</th>
<th>Free</th>
<th>Accy</th>
<th>Current</th>
<th>For</th>
<th>LnDep</th>
<th>InRec</th>
<th>LnAssets</th>
<th>AccyCorr</th>
<th>LnFEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 2: Correlation Matrix: SEP 200
be positive. All other coefficients have the expected signs, and are mostly statistically significant. I also performed an OLS regression of the underlying audit fee model,

\[
\text{LNFEE}_{i,t} = \kappa_{i,t} + \lambda_1 \times \text{LNASSETS}_{i,t} + \lambda_2 \times \text{OPINION}_{i,t} + \lambda_3 \times \text{INVREC}_{i,t} + \lambda_4 \times \text{LNEMP}_{i,t} + \lambda_5 \times \text{FOR}_{i,t} + \lambda_6 \times \text{CURRAT}_{i,t} + \eta_{i,t}.
\] (4.23)

The coefficient on INVREC remained positive in the German sample, which indicates that it is not the addition of the fitted AUDSCOR variable that causes troubles. Moreover, the adjusted \(R^2\)-values of 60.81\% (U.S.-sample) and 84.86\% (German sample) indicate that the underlying audit fee model performs decently.

<table>
<thead>
<tr>
<th>Table 3a: Regression Results S&amp;P 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>(2)</td>
</tr>
<tr>
<td>LNFEE 2SLS</td>
</tr>
<tr>
<td>Year-Fixed-Effects</td>
</tr>
<tr>
<td>AUDSCOR -0.508*** (-2.764)</td>
</tr>
<tr>
<td>LNASSETS 0.575*** (7.654)</td>
</tr>
<tr>
<td>OPINION -1.113*** (-6.402)</td>
</tr>
<tr>
<td>INVREC 1.182*** (2.694)</td>
</tr>
<tr>
<td>LNEMPLOYEES 0.146*** (2.706)</td>
</tr>
<tr>
<td>FOR 0.00671** (2.452)</td>
</tr>
<tr>
<td>CURRAT 0.0228 (0.508)</td>
</tr>
<tr>
<td>Constant 28.90*** (2.774)</td>
</tr>
<tr>
<td>Observations 2,877 2,877</td>
</tr>
</tbody>
</table>

Robust t-statistics in parentheses

*** \(p<0.01\), ** \(p<0.05\), * \(p<0.1\)
The goal of the empirical analysis is rather to get a feel for the validity of the proposed theory, than to provide a comprehensive check of its robustness. Apart from the statistical problem of missing data and the measurement error induced by using the ASSET4 ESG scores, the main limitation of the design is that it does not allow any inference about a causal effect. In other words, I do not directly test whether the substitution in the U.S.-sample, respectively the complementarity in the German sample are caused by equilibrium differences in the production technology $d(a, s)$, but only document an outcome that is consistent with the theory. However, as discussed at the end of the previous section, all supply and demand forces affect the equilibrium production technology in the sense that they change the costs and benefits of the $(s, a)$-investment. Hence, given that the strategic bond between the auditor and the client is sufficiently strong (which the worldwide use of a risk-based audit planning model suggests), the equilibrium resource allocation is likely to be driven by the behavior of the strategic effect $\frac{\partial a}{\partial s}$.

---

### Table 3b: Regression Results PRIME ALL SHARE

<table>
<thead>
<tr>
<th></th>
<th>(1) 2SLS</th>
<th>(2) 2SLS with Year-Fixed-Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNFEE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUDSCOR</td>
<td>0.0373**</td>
<td>0.0378**</td>
</tr>
<tr>
<td></td>
<td>(2.115)</td>
<td>(2.074)</td>
</tr>
<tr>
<td>LNASSETS</td>
<td>0.474***</td>
<td>0.478***</td>
</tr>
<tr>
<td></td>
<td>(8.147)</td>
<td>(8.098)</td>
</tr>
<tr>
<td>OPINION</td>
<td>-0.231</td>
<td>-0.100</td>
</tr>
<tr>
<td></td>
<td>(-0.754)</td>
<td>(-0.317)</td>
</tr>
<tr>
<td>INVREC</td>
<td>-0.325</td>
<td>-0.332</td>
</tr>
<tr>
<td></td>
<td>(-1.140)</td>
<td>(-1.151)</td>
</tr>
<tr>
<td>LNEMPLOYEES</td>
<td>0.147**</td>
<td>0.141**</td>
</tr>
<tr>
<td></td>
<td>(2.291)</td>
<td>(2.181)</td>
</tr>
<tr>
<td>FOR</td>
<td>0.00294**</td>
<td>0.00251</td>
</tr>
<tr>
<td></td>
<td>(2.003)</td>
<td>(1.641)</td>
</tr>
<tr>
<td>CURRAT</td>
<td>-0.110</td>
<td>-0.112</td>
</tr>
<tr>
<td></td>
<td>(-1.335)</td>
<td>(-1.301)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.968***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.728)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1,167</td>
<td>1,167</td>
</tr>
</tbody>
</table>

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1
4.7 Concluding Remarks

This paper provides a theoretical argument, and delivers empirical evidence on when firm-side governance activities, and external auditing are complements or substitutes. Anchoring on two shortcomings of prior research, it seeks to tie together theoretical and empirical research on the interplay of the institutions. Casually speaking, theoretical research in the field so far has been rigorous but narrow-minded, whereas the empirical picture is broader but rather fragmented. The analysis shows that two factors drive the equilibrium resource allocation. The first driver is the equilibrium strategic elasticity between the efforts of the firm and the auditor, that is either strategic complementarity or strategic substitution. The second characteristic is the interaction of this strategic elasticity with the environment - i.e. the strictness of enforcement mechanisms such as auditors’ third party liability.

Ultimately, the findings are intended to provide regulatory guidance, for example in the design of liability regimes. However, at least two major limitations of the model’s explanatory power should be pointed out. First, the economic and regulatory environment pictured is narrow. Although the regularity conditions can be relaxed without qualitatively changing the empirical predictions, nothing can be said about the equilibrium effects under different liability systems. For example, in the first dissertation paper on ICS quality and damage apportionment the two activities might appear to be complements in equilibrium, despite being constrained to be strategic substitutes. This is not possible in the present model. Second, the model variation with the similar-looking square root function shows how sensitive the results are to changes regarding the production technology $d(a,s)$. More research on the nature of the auditor-client collaboration is necessary to tailor suitable empirical models and derive real-world policy recommendations. The congruence of the model predictions with the empirical evidence however suggests that the exponential-like model specification might not be too far away from reality.
4.8 Appendix

**Derivation of the equilibrium audit effort**

\[
\frac{dU^A}{da} = 0 \iff s(1 - p)V e^{-sa} = k \iff \frac{s(1-p)V}{k} = e^{sa} \iff a^* = \ln \left[ \frac{s(1-p)V}{k} \right] / s.
\]

It remains to show that \( a \) is a maximum. The second derivative of (4.3) w.r.t. \( a \) reads

\[
\frac{d^2U^A}{da^2} = -s^2(1-p)V e^{-sa} < 0,
\]

which is negative. Hence \( a^* \) is a maximum of \( U^A \).

QED.

**Proof of Lemma 4.1**

Parts (i), (ii), and (iii) are trivial. To prove part (iv), differentiate the equilibrium effort as given in (4.6) w.r.t. \( s \). This gives

\[
\frac{da^*}{ds} = \frac{1 - \ln \left[ \frac{s(1-p)V}{k} \right]}{s^2},
\]

which is clearly positive (negative), if

\[
s < \frac{ek}{(1-p)V} \quad \text{or} \quad \frac{ek}{(1-p)V} < s.
\]

(4.24)

holds.

QED.
Derivation of the equilibrium audit fee

\[ U^A(a^*) \overset{!}{=} 0 \iff \]

\[ \iff W \left[ p + (1-p)(1 - \frac{k}{s^c(1-p)V}) + (1-p)\frac{k}{s^c(1-p)V} \right] \overset{!}{=} \]

\[ \iff W = \frac{k}{s^c} \ln \left[ \frac{s^c(1-p)V}{k} \right] \left[ p + (1-p)(1 - \frac{k}{s^c(1-p)V}) + (1-p)\frac{k}{s^c(1-p)V} \right] + V \left( 1-p \right) \frac{k}{s^c(1-p)V} \iff \]

\[ \iff W = \frac{k}{s^c} \ln \left[ \frac{s^c(1-p)V}{k} \right] \left[ p + (1-p)(1 - \frac{k}{s^c(1-p)V}) + (1-p)\frac{k}{s^c(1-p)V} \right] + V \left( 1-p \right) \frac{k}{s^c(1-p)V} \iff \]

\[ \iff W = \frac{k}{s^c} \left\{ \ln \left[ \frac{s^c(1-p)V}{k} \right] + 1 \right\}. \]  

(4.26)

Proof of Remark 4.1

First, rewrite the equilibrium cost of capital as

\[ \alpha^* = e^{-sa}(1-p)(I-V) + pI \iff \]

\[ \iff \frac{k}{s^c v}(1-p)(I-V) + pI \iff \frac{k}{s^c v}(I-V) + pI. \]

Inserting this term and the audit fee expression (4.25) into (4.1) yields

\[ U^A(a^*, \alpha^*) = p(1 - \frac{k}{s^c v}(I-V) + pI) Z - \frac{k}{s^c v} \left\{ \ln \left[ \frac{s^c(1-p)V}{k} \right] + 1 \right\} - c(s) \iff \]

\[ \iff pZ - pI - \frac{k}{s^c v}(I-V) Z - W(s^c) - c(s) \iff \]

\[ \iff pZ - pI - \frac{k}{s^c v}(I-V) - W(s^c) - c(s). \]

Since the fee \( W \) is fixed based on the conjecture \( s^c \) before the auditor is hired, it does
not influence the firm’s decision at the first stage. Differentiating (4.26) w.r.t. \( s \) gives the FOC

\[
\frac{dU^C(a^*, \alpha^*)}{ds} = k \frac{(I - V)}{V} - c'(s) = 0. \tag{4.27}
\]

In equilibrium, the second derivative of \( U^C \) w.r.t. \( s \) reads

\[
\frac{d^2U^C}{ds^2} = \frac{k}{(s^*)^3} \left[ -2 \frac{(I - V)}{V} \right] - c''(s),
\]

which is clearly negative for all \( s^* \). Since the production function \( d(a, s) \) already exhibits decreasing returns it is clear that convexity of the cost-function \( c(s) \) is not even a necessary condition for the second derivative to be negative.\(^{164}\)

QED.

**Proof of Proposition 4.1**

In equilibrium, \( \frac{d^2U^C}{dsdV} \) must hold. Implicit differentiation yields

\[
\frac{d^2U^C}{dsdV} = \frac{\partial^2U^C}{dsdV} + \frac{d^2U^C}{ds^2} \frac{ds^*}{dV} = 0 \iff \\
\iff \frac{ds^*}{dV} = -\frac{\frac{\partial^2U^C}{dsdV}}{\frac{d^2U^C}{ds^2}}.
\]

Since the denominator is negative due to the strict concavity of \( U^C \), the sign of \( \frac{ds^*}{dV} \) is equal to the sign of the partial derivative \( \frac{\partial^2U^C}{dsdV} \). This derivative reads

\(^{164}\) In fact, even a (locally) slightly concave cost function would suffice. Such a cost-function would capture learning effects of the auditor.
\[
\frac{d^2 U^C}{ds dV} = - \frac{k(I - V)}{(s^*)^2 V^2} - \frac{k}{(s^*)^2 V^2} \Leftrightarrow \\
\Leftrightarrow - \frac{kI}{(s^*)^2 V^2} < 0,
\]

and is negative for \( V < I \).

QED.

**Proof of Proposition 4.3**

The auditor’s FOC reads

\[
\frac{dU^A}{da} = 0 \Leftrightarrow se^{-(sa)^*}(1 - p)V - ka^* \frac{1}{V} = 0.
\]

In equilibrium,

\[
\frac{d^2 U^A}{dadV} = 0 \Leftrightarrow \frac{\partial^2 U^A}{da \partial V} + \frac{\partial^2 U^A}{da^* \partial s} dV + \frac{\partial^2 U^A}{da} \frac{da^*}{dV} \Rightarrow \\
\Leftrightarrow \frac{da}{dV} = - \frac{\frac{\partial^2 U^A}{da \partial V} + \frac{\partial^2 U^A}{da^* \partial s} dV}{\frac{\partial^2 U^A}{da^2}}
\]

must hold. Due to the strict concavity of \( U^A \), the numerator determines the sign of \( \frac{da}{dV} \).

Let the numerator be denoted by \( N \). The full form of \( N \) reads

\[
N = s^* e^{-(sa)^*}(1 - p) + \left[ e^{-(sa)^*}(1 - p)V - (sa)^* e^{-(sa)^*}(1 - p)V \right] \frac{ds^*}{dV} \Leftrightarrow \\
\Leftrightarrow e^{-(sa)^*}(1 - p) \left[ s^* - V \frac{kI}{(s^*)^2 V^2} (1 - (sa)^*) \frac{1}{-\frac{d^2 U^C}{ds^2}} \right].
\]

193
Since the factor $e^{-(sa)^*}(1 - p)$ is strictly positive, the sign of $N$ and thus the sign of $\frac{da^*}{dV}$ is equal to the sign of

$$s^* \frac{k}{(s^*)^3} \left[ 2 \frac{(I - V)}{V} \right] + \frac{k}{(s^*)^2} (1 - (sa)^*) \frac{-I}{V} \leftrightarrow \frac{k}{(s^*)V} [2(I - V) - (1 - (sa)^*) I] \leftrightarrow \frac{k}{(s^*)^2 V} \left\{ 2(I - V) - \left[ 1 - \ln \left( \frac{s(1-p)V}{k} \right) \right] I \right\}. \quad (4.28)$$

Now note from the derivation of $a^*$ and Lemma 4.1 that the sign of $(1 - (sa)^*)$ is equal to the sign of $\frac{\partial a^*}{\partial s}$. This means, if $(1 - (sa)^*) < 0$, $s^*$ and $a^*$ are strategic substitutes (complements). Under strategic substitution, (4.28) is clearly positive. However, under equilibrium strategic complementarity the sign is ambiguous. It is obvious that the first addend of the parenthetical term in (4.28) does not depend on the auditor’s effort choice. Therefore, if the auditor chooses an effort $a^*$, such that the equilibrium strategic complementarity is sufficiently strong, that is the term $(1 - (sa)^*)$ becomes sufficiently large, the sign of (4.28), and thus the sign of $\frac{da^*}{dV}$ becomes negative.

QED.

Proof of Remark 4.2

The client firm’s modified utility function reads

$$U_{mod}^C = p(1 - \alpha)Z - W(s^c) - cs + ra - F(r).$$

Since $r$ has no effect on the expected utilities of the investors and the auditor, the expressions for the equilibrium return share $\alpha^*$ and the equilibrium audit quality $a^*$ are given
by (4.4) and (4.6). After choosing the level of internal governance $s$, the firm purchases NAS $r$ from the auditor. Since the auditor must break even in equilibrium, the fee $F(r)$ is equal to the auditor’s NAS-effort costs $r^2/2$. Differentiating (4.19) w.r.t. $r$, and solving the condition $\frac{dU_{\text{mod}}^c}{dr} = 0$ for $r$ gives

$$r^* = a^*,$$

which is a maximum due to $\frac{d^2U_{\text{mod}}^c}{dr^2} = -1$. Inserting $r^* = a^*$ into (4.19), and differentiating (4.19) w.r.t. $s$ gives the modified FOC for $s$

$$\frac{dU_{\text{mod}}^c}{ds} = \frac{k \left(I - V\right)}{s^2 V} - c + a^* \frac{\partial a^*}{\partial s} = 0. \quad (4.29)$$

Recall that $a^* \geq 0$ requires

$$k \leq s(1-p)V \Leftrightarrow s \geq \frac{k}{(1-p)V}.$$

Let $\frac{k}{(1-p)V}$ be denoted by $s^*$. For $s = s^*$, $a^*$ is zero. Therefore (4.29) is positive, whenever $c \leq \bar{c} = \frac{(1-p)^2 \left(I - V\right)}{kV}$. Conversely, for $s \to \infty$ we have

$$\lim_{s \to \infty} a^* = \frac{\ln \left[ \frac{\infty(1-p)V}{k} \right]}{\infty} = \frac{\infty}{\infty},$$

and thus also

$$\lim_{s \to \infty} \frac{\partial a^*}{\partial s} = 1 - \frac{\ln \left[ \frac{\infty(1-p)V}{k} \right]}{\infty} = -\frac{\infty}{\infty}.$$

Application of L’Hospital’s rule however gives

195
\[ \lim_{s \to \infty} a^* = \lim_{s \to \infty} \frac{1}{s} = 0, \]

and

\[ \lim_{s \to \infty} \frac{\partial a^*}{\partial s} = \lim_{s \to \infty} -\frac{1}{2s^2} = 0. \]

Therefore, (4.29) becomes negative for \( s \to \infty \) for all \( c > 0 \). The existence of a value \( s^* \) that solves (4.29) then follows from the intermediate value theorem. Now, take the second derivative

\[ \frac{d^2 U_{\text{mod}}}{d^2 s} = -\frac{2k (I - V)}{s^3} + \left( \frac{\partial a^*}{\partial s} \right)^2 + a^* \frac{\partial^2 a^*}{\partial s^2} \leftrightarrow -\frac{2k (I - V)}{s^3} + \frac{1 + sa^* (-5 + 3sa^*)}{s^4} \leftrightarrow \]

\[ \leftrightarrow \frac{1}{s^3} \left\{ \frac{-2k (I - V)}{V} + \frac{1 + sa^* (-5 + 3sa^*)}{s} \right\} \rightarrow P. \]

It is obvious that the sign of the second derivative is given by the sign of the parenthetical term \( P \). First, consider the case \( s^* = \bar{s} \). Since \( \bar{s} > 0 \) but \( a^* \) is zero at \( \bar{s} \), the parenthetical term reduces to

\[ -\frac{2k (I - V)}{V} + \frac{(1 - p) V}{k}. \]

This term is clearly negative for sufficiently high \( I \). Second, consider the case \( s \to \infty \).

Since \( \lim_{s \to \infty} sa^* = \lim_{s \to \infty} \ln \left[ \frac{s(1-p)V}{k} \right] = \infty \), the limit \( \lim_{s \to \infty} P \) reads
\[
\lim_{s \to \infty} P = \left\{ -2k \frac{(I-V)}{V} + \lim_{s \to \infty} \frac{1 + sa^*(-5 + 3sa^*)}{s} \right\} \Leftrightarrow \\
\Leftrightarrow -2k \frac{(I-V)}{V} + \frac{\infty}{\infty}.
\]

L'Hospital's rule gives the limit of the second addend, which is

\[
\lim_{s \to \infty} \frac{1 + sa^*(-5 + 3sa^*)}{s} = \lim_{s \to \infty} \frac{1}{s} (-5 + 3sa^*) + \frac{3sa^*}{s} = \lim_{s \to \infty} \frac{-5}{s} + 6a^* = 0.
\]

Therefore we have \(\lim_{s \to \infty} P = -2k \frac{(I-V)}{V} < 0\). It follows from continuity, that for sufficiently high \(s = \infty - \epsilon\), whereas \(\epsilon\) is a small arbitrary number, the second derivative \((4.30)\) is negative.

Given that the second derivatives are locally negative for sufficiently low and sufficiently high \(s\) (= the regions where equilibrium strategic complementarity, respectively strategic substitution prevail), the sign of the effect \(\frac{da^*}{dV}\) is given by the sign of

\[
\frac{d^2 UC}{dsdV} = -\frac{k(I-V)}{(s^*)^2V^2} - \frac{k}{(s^*)^2V} + \frac{\partial a^*}{\partial V} \frac{\partial a^*}{\partial s} + a^* \frac{\partial^2 a^*}{\partial s \partial V} \Leftrightarrow \\
\Leftrightarrow -\frac{kl}{(s^*)^2V^2} + \frac{1}{s^*V} \left[ \frac{\partial a^*}{\partial s} - \frac{a^*}{V} \right].
\]

It is obvious that the whole expression \((4.32)\) is negative if \(\frac{\partial a^*}{\partial s} < 0\). In this case \(\frac{ds^*}{dV}\) is negative, but

\[
\frac{da^*}{dV} = \frac{\partial a^*}{\partial V} > 0 + \frac{\partial a^*}{\partial s} \frac{ds^*}{dV} < 0
\]

is positive. Hence, the observable equilibrium relation is substitution. Conversely, we have
\[
\lim_{s \to 2} \frac{d^2U^C}{dsdV} = -\frac{I(1-p)^2}{k} + \frac{(1-p)^3V^2}{k^3} = \frac{(1-p)^2}{k^3}(-k^2I + (1-p)V^2), \quad (4.33)
\]

which is positive for sufficiently small \(k\). It follows from continuity, that (i) for \(I\) that are sufficiently large, such that \(\frac{d^2U^C}{dsdV}\) is negative for small \(s^*\), and (ii) for \(k\) that are sufficiently small, such that \(\frac{d^2U^C}{dsdV}\) is positive for small \(s^*\), both \(\frac{ds^*}{dV}\) and

\[
\frac{da^*}{dV} = \frac{\partial a^*}{\partial V} + \frac{\partial a^*}{\partial s} \frac{ds^*}{dV} > 0 > 0
\]

are positive.\(^{165}\) Hence, the observable equilibrium relation is complementarity.

Now turn to the equilibrium fees \(W(s^c)\) and \(F(r^*)\). In equilibrium, \(W(s^c) = W(s^*)\) holds, and the total audit fees are set such that the auditor just breaks even. Therefore we have

\[
W(s^c) = ka^* + (1-p)e^{-(sa)^*}V \Leftrightarrow
\]

\[
\Leftrightarrow \frac{ln\left[\frac{s^*(1-p)V}{k}\right]}{s^*} + \frac{k}{s^*} \Leftrightarrow
\]

\[
\Leftrightarrow \frac{k}{s^*} \left[1 + ln\left[\frac{s^*(1-p)V}{k}\right]\right],
\]

and

\[
F(r^*) = \frac{a^2}{2}.
\]

Let the aggregate fee \(W(s^c) + F(r)\) be denoted by \(G(s)\). Differentiating \(G(s)\) w.r.t. \(s\) gives

\(^{165}\) It can be easily seen from (4.31) and (4.33) that the model offers enough degrees of freedom for both conditions to be simultaneously satisfied.
The first part $\frac{dW(s)}{ds}$ is clearly negative. For $\frac{\partial a^*}{\partial s} < 0$, the second part is also negative, that is the aggregate audit fee $G(s)$ decreases in $s$. Now, rewrite

$$\frac{dG(s)}{ds} = \frac{dW(s)}{ds} + a^* \frac{\partial a^*}{\partial s} \Leftrightarrow -\frac{k}{s^2} \left\{ \ln \left[ \frac{s^*(1-p)V}{k} \right] \right\} + a^* \frac{\partial a^*}{\partial s}.$$ (4.34)

To investigate the behavior of $G(s)$ for sufficiently low $s$, insert $s = \bar{s}$ into the parenthetical term, which gives

$$-(1-p)V + \frac{(1-p)^2V^2}{k^2} \Leftrightarrow (1-p)V \left[ -1 + \frac{(1-p)V}{k^2} \right].$$ (4.35)

This term is positive for sufficiently small $k$.\footnote{More precisely, for $k$ that are small enough such that (4.33) is positive, this term is also positive, given that $I > 1$. An initial investment $I > 1$ is likely, as (4.31) requires a sufficiently large $I$ to be positive. Conversely, if $I < 1$, then any $k$ that is sufficiently small such that (4.35) is positive, also implies that (4.33) is positive.} It follows from continuity, that within a range of sufficiently small $s > \bar{s}$, $G(s)$ increases in $s$.

QED.
5 Conclusion

5.1 On the Importance of Counter-Intuitive Results

The attentive reader of this dissertation will have noticed the emphasis placed on "counter-intuitive" findings, that is results which go against a reasonable conjecture or a conventional wisdom. There is a common sense in theoretical research that these are the most intriguing results. It is needless to say that it is not the goal of theoretical economic research to produce a mass of counter-intuitive results by tailoring suitable models. The model itself is only an instrument that subjects thought processes to logical rigor. If the model is appropriately specified, it can explain virtually anything, regardless of whether the phenomenon can be observed in reality or not. However, aiming for the unexpected is essential in social sciences research, not only because it is the expression of scientific curiosity. It is the rigorous scrutiny of what we believe to understand, as well as the pursuit of insight into phenomena we do not yet understand. Even more importantly, models that are designed to produce counter-intuitive findings force a deep thinking about the underlying economic mechanics. Often the counter-intuitive results only hold under certain conditions, and most of the time the search for the counter-intuitive finding ends with the insight that the world is more intuitive than the mass of working papers and journal publications suggests. However, it is the thought processes leading to these (often disappointing) insights that promote the understanding of social interaction. That said, the answer to the main research question of the dissertation as outlined in the in-

\[167\] This is famously illustrated by Postrel (1991), who gives a logical argument why it can be rational for a bank manager to set his trousers on fire on the open street in order to signal the quality of his bank.
CONCLUSION

...roduction, is not primarily to be found in the counter-intuitive propositions of the three papers. Whether these or their intuitive opposites hold in reality is an empirical question. The answer is rather to be found in the lines of thought in between the unexpected comparative-static results. In that spirit, this final chapter spans the distance between the three individual papers by summarizing the economic forces that shape both the intuitive and counter-intuitive aspects of strategic auditor-client interaction. The chapter also gives an outlook to potential future research in the field.

For the sake of readability, I abbreviate the paper The Impact of Damage Apportionment on ICS Quality and Financial Reporting Accuracy as ICS-paper, the paper Audit Quality, Legal Liability, and the Audit Market under Risk Aversion as risk aversion-paper, and the paper Sequential Auditor-Client Interaction under Strategic Effort Complementarity as effort complementarity-paper hereafter.

5.2 The General Auditor-Client Model

The strategic auditor-client relationship with regard to the sequential joint production of accurate accounting information is described by the following general model:

\[
\max_s \ E(U^C(s)) = pU^C(\Pi(a^*, s) - c(s) - W(a^*, s)) + (1 - p)d(a^*, s)U^C(\Pi(s) - W(a^*, s)) + (1 - p)(1 - d(a^*, s))U^C(\Pi_V(s) - W(a^*, s)) \\
\text{s.t.} \quad a^* \in \arg\max_a E[U^A(a | s)] = [p + (1 - p)d(a, s)]U^A(W - k(a)) + (1 - p)(1 - d(a, s))U^A(W - V_A - k(a)) \\
\text{s.t.} \quad E(U^A(a^*, s^*)) \geq U^A \\
\text{s.t.} \quad E(U^C(a^*, s^*)) \geq U^C
\]
Although none of the three dissertation papers uses the maximum configuration of the model, each of its elements have yet been extensively described and discussed. I therefore pare the model description down to the minimum, but focus on the big picture. $\mathbb{E}(U^C)$ and $\mathbb{E}(U^A)$ are the client firm’s and the auditor’s VNM expected utility functions, whereas $U^C$ and $U^A$ are their Bernoulli utility functions. $U^C$ and $U^A$ mark their reservation utilities, that is the expected utility each of them draws out of his or her second-best alternative.

The variable $s$ depicts the firm-side activities to make financial reports more accurate (internal controls, audit committee activity etc.), whereas $a$ marks the auditor’s choice of audit effort. The effort $s$ is chosen before the auditor chooses the effort $a$. Importantly, the auditor can observe $s$ prior to choosing $a$. The function $d(a, s)$ translates the effort investments $s$ and $a$ into an improvement in financial reporting accuracy, as reflected by a decrease in the error rate from initially $(1 - p)$ to $(1 - p)(1 - d(a, s))$. $\Pi$ is the shareholders’ (and managers’) additional consumption made possible by the positive capital market reaction following an accurate positive report. It is the incentive to spend resources into the production of accurate financial reports. Depending on the ability of outside investors to observe the firm’s and the auditor’s investments into accurate reporting, and their financial consequences in case of misreporting, $\Pi$ may or may not be dependent on the efforts $s$ and $a$. The functions $c(s)$ and $k(a)$ are the direct costs caused by the respective efforts. $V_A$ and $V_C$ depict the auditor’s and the firm’s legal liability consequences following the publication of an erroneous report, and are the parameters of primary interest. These may be dependent on each other. Finally, $W$ depicts the audit fee. To preserve the independence of the auditor’s professional judgment, the fee has to be non-contingent on any outcome of the audit. Therefore, from the auditor’s point of view the fee is fixed at the time he or she decides on the extent of the audit procedures. From the client firm’s perspective, the fee depends on the anticipated audit effort, since it has to compensate the auditor for his or her costs. Depending on whether the auditor can observe the firm’s effort choice before the fee is fixed, the firm may also consider an effect of $s$ on the fee.
The structure of the model is common knowledge to all players.

There are two types of information asymmetry captured in this model. First, the reporting firm is initially uninformed about its own type, respectively the quality of new investments into the firm, as represented by financial statement figures. Conceptually, this is an adverse selection problem between the firm and some unmodeled outside party. This adverse selection problem is the reason why the firm engages in information-producing effort $s$, and hires an external auditor. Second, there is a moral hazard problem because the goals of the reporting firm and the auditor deviate. The firm wants to optimize the aggregate information production in order to obtain a maximal benefit $\Pi$ at the minimum aggregate costs, whereas the auditor does not care about $\Pi$, but only about minimizing his or her total costs. The moral hazard problem arises, because the firm cannot observe and enforce the desired audit effort. As outcome-contingent audit fees are prohibited, the typical solution to moral hazard in principal-agent relationships, which is to reward and punish the agent based on the outcome of his or her effort, is not applicable. Hence, the firm has to use its effort $s$ not only to generate the desired information, but also to provide incentives to the auditor to act in his or her best interest.

From a purely technical perspective, the model depicted above is a standard principal-agent model in which the incentives to the auditor (the agent) are set by the firm (the principal) in the sense that his or her effort determines the marginal output rate of the auditor’s effort, primarily but not necessarily exclusively through its effect on the error rate in financial statements. The standard approach to solve the program (5.1)-(5.4) is the first-order approach (FOA). The FOA involves replacing the continuum of the auditor’s incentive compatibility constraints (5.2) by the auditor’s FOC $\frac{dE(U^A(a|s))}{da} \neq 0$. As famously shown by Mirrlees (1975) and Rogerson (1985), the FOA is generally not valid because the FOC only requires the agent’s expected utility to be stationary. However, a stationary point is not necessarily a global maximum. A sufficient condition for the FOA to be valid is the strict concavity of the agent’s expected utility in his or her effort. This ensures that any
CONCLUSION

stationary point is a global maximum. There are basically two ways to make the auditor’s expected utility function concave. First and very realistically, one can assume that the audit technology is either decreasingly productive, or increasingly costly at the margin. Both characteristics imply decreasing returns to scale. The two commonly used sufficient conditions for this to be true are the monotone likelihood-ratio condition (MLRC), and the convexity of the distribution function condition (CDFC). These conditions are referred to as the Mirrlees-Rogerson conditions. As shown by Rogerson (1985), together with a set of eight regularity conditions, the MLRC and the CDFC are sufficient for concavity of the agent’s expected utility in his or her effort. Second, the concavity requirements can also be loaded onto the agent’s preference structure, that the Bernoulli utility curve. A set of necessary and sufficient conditions for the agent’s decision problem to be concave without imposing the CDFC is given by Jewitt (1988).

In all three dissertation papers, the auditor’s expected utility function is strictly concave. In the ICS-paper and the effort complementarity-paper the concavity requirements are imposed on the audit technology. The model in the risk aversion-paper does not specify the auditor’s preference structure and technology. Therefore the source of the concavity is left open. Because the auditor chooses \( a \) after the firm chooses \( s \), the firm optimizes over its expected utility \( E(U^C(s)) \), anticipating the equilibrium effort choice \( a^* \) of the auditor. The firm’s equilibrium expected utility function does not necessarily have to be (strictly) concave for the constrained maximization program (5.1)-(5.4) to be solved. However, since that the firm and the auditor jointly produce the same information, it is reasonable to assume that also the effort \( s \) exhibits decreasing returns to scale. Given that the decrease in the marginal output of \( s \) per resource input is sufficiently strong, the client firm’s expected utility function (5.1) is strictly concave in \( s \). This assumption is technically convenient because it avoids the detailed examination of potential corner solutions, and facilitates the comparative static analysis. Hence, in the ICS-paper and the effort complementarity-paper, where \( s \) is a continuous choice variable, the firm’s equi-
librium effort-choice $s^*$ is given by the FOC \( \frac{d \mathbb{E}(U_C(s))}{ds} \bigg| s = 0 \). In the risk aversion-paper the firm’s effort choice is not explicitly modeled.

Conditions (5.3) and (5.4) are the auditor’s and the client firm’s participation constraints. Condition (5.3) ensures that the auditor is willing to accept the engagement, whereas condition (5.4) ensures that the client wants to hire an auditor at all. These two conditions depict the distribution of bargaining power between the auditor and the client, and thus implicitly carry assumptions about the structure of the audit market. Most prior papers assume that there is perfect price competition on the audit market. This means that no auditor earns a rent if hired, and the auditor’s participation constraint is thus always binding. Any surplus that is generated by the audit is then solely collected by the client firm. I follow this widely accepted and empirically supported (see London Economics/ Ewert (2006)) assumption in the ICS-paper and the effort complementarity-paper. In the risk aversion-paper, where a main focus is placed on the effects of liability regulation on the audit market structure, a richer modeling of the bargaining power is chosen. Following the model of Dye (1995), the distribution of the surplus generated by the audit depends on the relation between the supply and the demand for audit services. Another assumption that is implicitly captured in (5.3) is the assumption that each auditor has the same reservation utility. This assumption is a simplifying normalization that allows the representation of (infinitely) many auditors’ participation decisions in one condition. It is also realistic to assume that each auditor’s second-best alternative to run an audit firm is to become an employee of an audit firm at a market-conform fixed salary.

Given the applicability of the first-order approach, the solution of the model is given by the FOCs

\[
\frac{d \mathbb{E}(U^A(a \mid s))}{da} \bigg| a = 0,
\]

and

205
\[
\frac{d\mathbb{E}(U^C(s))}{ds} \neq 0 \iff \frac{\partial\mathbb{E}(U^C(s))}{\partial s} + \frac{\partial\mathbb{E}(U^C(s))}{\partial a} \frac{\partial a^*}{\partial s} = 0.
\]

From the ICS- and the effort complementarity-paper we know that the first addend in the second equation represents the firm’s incentive to implement accurate reporting on its own, whereas the second addend represents the incentive to steer the auditor’s effort via the choice of \(s\).

By the implicit function theorem, the first-stage-change in the equilibrium effort investment of the firm \(s^*\) after some exogenous (regulatory) shock \(d(\cdot)\) is given by

\[
\frac{\partial^2\mathbb{E}(U)}{ds \partial d(\cdot)} = \frac{\partial^2\mathbb{E}(U)}{\partial s \partial d(\cdot)} + \frac{\partial^2\mathbb{E}(U)}{\partial a \partial d(\cdot)} \frac{\partial a^*}{\partial s} + \left[ \frac{\partial^2\mathbb{E}(U)}{\partial s \partial a} + \frac{\partial^2\mathbb{E}(U)}{\partial a \partial s} \right] \frac{\partial a^*}{\partial s} + \frac{\partial\mathbb{E}(U)}{\partial a} \frac{\partial^2 a}{\partial s \partial d(\cdot)}.
\]

(5.5)

The second-stage-change in the equilibrium effort invested by the auditor is given by

\[
\frac{\partial^2\mathbb{E}(U)}{da \partial d(\cdot)} = \frac{\partial^2\mathbb{E}(U)}{\partial a \partial d(\cdot)} + \frac{\partial^2\mathbb{E}(U)}{\partial a \partial s} \frac{ds^*}{d(\cdot)},
\]

(5.6)

whereas it is more convenient to write the second-stage-change explicitly as

\[
\frac{da^*}{d(\cdot)} = \frac{\partial a^*}{d(\cdot)} + \frac{\partial a^*}{ds} \frac{ds^*}{d(\cdot)}.
\]

(5.7)

Each and every result in this dissertation that regards the equilibrium effort choices \(a^*\) and \(s^*\) boils down to the equations (5.5) and (5.7). The underlying economic forces are either represented by separate partial derivatives, or a mesh of partial derivatives. The
following classification of these forces shall not only interconnect the results of the three papers, but also serve as a guideline for future, empirically oriented modeling.

5.3 First-Order Effects

A first-order effect is a direct causal effect of $A$ on $B$. An argument based on a counter-intuitive first-order effect is the most immediate explanation for an unexpected theoretical prediction, or an unexplained empirical phenomenon. These arguments are typically compelling and robust, because they require very little structure, such as assumptions about the strategic interaction of the decision-maker with other parties. At the same time, new and unexpected first-order effects offer an enormous research potential, since they trigger a whole lot of higher-order effects. This may put prior findings, or - depending on how fundamental the first-order effect is - a whole research strand into a new perspective. On the downside, counter-intuitive first-order effects are hard to detect. By definition, first-order effects are the most obvious causal relations. Our understanding of what is economically intuitive, or which behavior is rational, is based on the behavior we observe in reality the most often. Hence, the amount of "un-intuitiveness" we can observe, and for which we can search scientific explanations is limited. However, if we encounter an unexplained phenomenon, the natural first step to explain it is to search for first-order effects. Therefore, the counter-intuitive first-order effects are rapidly grazed by researchers, and - provided that the underlying premises of the provided explanations are reasonable - rather quickly become intuitive.

In the auditor-client model used in this dissertation, counter-intuitive first-order effects of exogenous factors on the auditor’s and client firm’s equilibrium behavior are formally represented by the partial derivatives
CONCLUSION

\[
\frac{\partial^2 E(U^C)}{d\sigma \partial(\cdot)} = \frac{\partial^2 E(U^C)}{d\sigma \partial(\cdot)} + \frac{\partial^2 E(U^C) \partial a^*}{d\sigma} + \left[ \frac{\partial^2 E(U^C) \partial a^*}{d\sigma} + \frac{\partial^2 E(U^C) \partial a^*}{d\sigma} \right] \frac{\partial a^*}{\partial(\cdot)} + \frac{\partial E(U^C) \partial a^*}{d\sigma} \frac{\partial a^*}{\partial d(\cdot)}. 
\]

and

\[
\frac{da^*}{d(\cdot)} = \frac{\partial a^*}{d\sigma} + \frac{\partial a^* ds^*}{d\sigma d(\cdot)}. 
\]

Of the analytical models in the three dissertation papers, the one in the risk aversion-paper is developed around an ambiguous, partly counter-intuitive first-order effect. Intuitively, one would expect that more risk-averse auditors audit more carefully in order to eliminate the liability risk. However, it turns out that also the opposite can be true. This result is per se not new, as it has been shown and discussed in the general decision theory literature. However, with regard to (theoretical) auditing research, the result is highly counter-intuitive, since it puts the legitimacy of the ubiquitous simplifying assumption of auditors being risk-neutral into question. The main contribution of the paper is that it shows how the negative effect of risk aversion on audit effort can propagate into negative direct effects of legal liability regulation on audit effort, and from there on to the market-supply and -demand of auditing services. In particular, the result that an increase in the absolute liability consequences can directly cause a lower audit effort choice is highly counter-intuitive in the context of existing theoretical auditing research.

I believe that counter-intuitive first-order effects offer a huge potential for future research. Unintuitive and seemingly irrational decisions are a main research subject of (mathematical) decision theory. While the accounting and auditing research community is certainly not aware of all the surprising effects discovered by general decision theory, mathematical economists seem to care little about accounting and auditing. I anticipate major contributions at three consecutive stages: First, the counter-intuitive first-order effects are

208
to be transferred from general decision theory into the world of accounting and auditing by investigating their interaction with institutional features such as auditor liability. Second, the implications of these effects for the strategic interaction between the auditor, the reporting firm and the capital market have to be assessed. Third and finally, one can study how the equilibrium changes if more than just one party is subject to the counter-intuitive first-order effect. For example, it could be (realistically) assumed that also the client firm is risk-averse. The question is, whether the additional insights of such a combined analysis are worth overcoming the inevitable mathematical challenges.

5.4 Second-Order Effects

Although the majority of papers (including mine) focus on the main pillars of legal liability systems as trigger effects, these are per se not the reason behind most of the unexpected comparative static results derived from economic models. If that were indeed the case, the fundamental economic forces of legal liability would be largely unknown. However, given our view of the auditing process and our idea of human decision-making, we seem to have a rather solid understanding of these basic forces. For example, it is generally accepted (and very easy to show with a standard decision model) that the direct effect of higher legal liability consequences on the effort supplied by a risk-neutral auditor is positive. Yet, there are numerous model-theoretic papers showing a negative causal effect of the liability consequences on the audit effort supplied by a rational, risk-neutral auditor. The reason for the unexpected direction of these effects is that the positive first-order effect is outweighed by one or more negative second-order effects. For example, higher

---

168 A good example is Bigus (2012), who investigates how ambiguity aversion, that is the aversion to uncertainty about the second moment of a probability distribution (famously documented by Ellsberg (1961)), interacts with vaguely defined auditing standards. At first glance, the argument of Bigus (2012) appears quite similar to mine. However, in Bigus (2012) the counter-intuitive negative effect of the liability amount on audit effort stems from the assumption that the auditor perceives the liability risk to be to some degree invariant in his or her effort, although this is not the case. This assumption is not compatible with expected utility theory. In my paper, the negative effect solely stems from the fact that a risk-averse decision maker’s Bernoulli utility curve is concave.
liability consequences directly increase a risk-neutral auditor’s incentive to exert high effort. However, they may also increase the marginal efficiency of internal controls. In this case, the firm would have an incentive to invest more into internal controls, which in turn would lower the marginal efficiency of auditing. If this negative second-order effect is strong enough, it outweighs the positive first-order effect. Then the aggregate effect of higher liability on audit effort is negative. I am careful to use the subjunctive form here, because the proposed effect chain is anything but general. The direction and the magnitude of the second-order effects in this type of model depend on several factors.

5.4.1 Extent, Composition, and Distribution of Liability Incentives

The basic function of auditors’ third party liability is to insure investors against the losses from bad investments that were based on misstated financial reports. This insurance ex ante reduces the risk of providing funds to the reporting firm. Depending on the distribution of bargaining power between the firm and its potential investors, this risk reduction is reflected in a lower cost of capital. Under the reasonable assumption that the capital market is perfectly competitive, any expected benefit from the damage compensation is fully passed on to the firm via the cost of capital. However, there is no “free lunch” for the firm. In order to get an audit and thereby enable the cost of capital reduction, it has to compensate the auditor for his or her expected liability. It follows that the expected wealth change due to the damage payments is null for the auditor, the client firm, and the investors. This means that any change in the damage payments ceteris paribus leaves the expected utility of all three parties unchanged. Following Laux / Newman (2010), this mechanism is called ”triangle effect”. It is important to note that in general the triangle effect only holds under ubiquitous risk neutrality. If one or more parties are risk averse, the utility loss of the auditor from paying damages is unlikely to equal the investors’ utility gain from receiving damages. Therefore, the part of the audit fee that compensates the auditor for his or her expected liability is unlikely to equal the cost of
capital reduction due to the investors’ insurance through the damages. To account for these effects, regulators must take into account the unobservable individual risk preferences of auditors, firms and investors. This is not completely impossible, but requires to make assumptions about the connection between the risk preference and observable attributes, such as wealth or firm size. The beauty of the triangle effect is that it blanks out these largely unknown utility effects of the wealth transfer, and instead draws the attention to incentive effects that are independent of the individual risk attitude and thus much easier to steer for a regulator.

While the direct utility consequences of changes in the damage payments are null given that the triangle effect holds, changes in the damages nevertheless alter the incentives of the auditor. This is because the audit fee has to be non-contingent on the audit outcome, and is thus fixed at the time the auditor chooses his or her effort \( a \). Therefore, higher damages increase the equilibrium \( a^* \). However, the critical question is whether the fee is yet fixed at the time the firm chooses its effort \( s \). This depends on whether the auditor can observe the equilibrium \( s^* \) before the fee is fixed. If he or she can, then the direct incentive effect of the liability change on the firm’s effort choice, as given by

\[
\frac{\partial^2 E(U_C)}{ds \partial (\cdot)} = \frac{\partial^2 E(U_C)}{\partial a \partial (\cdot)} \frac{\partial a^*}{\partial s} + \left[ \frac{\partial^2 E(U_C)}{\partial s \partial a} + \frac{\partial^2 E(U_C)}{\partial a^* \partial s} \right] \frac{\partial a^*}{\partial (\cdot)} + \frac{\partial E(U_C)}{\partial a} \frac{\partial^2 a^*}{\partial s \partial (\cdot)},
\]

with \( d(\cdot) \) being an increase of the damages paid to investors, is null. This is true because the incentive to cut down \( s \) due to the improved investor insurance is exactly offset by the incentive to increase \( s \) in order to limit the auditor’s liability threat and thereby reduce the audit fee. Conversely, if the auditor cannot observe \( s^* \) at the time the fee is fixed, then the effect \( \frac{\partial^2 E(U_C)}{ds \partial (\cdot)} \) is negative. In this case, only the benefits of improved investor insurance provide direct effort-incentives to the firm, while the costs do not. Most prior papers, including [Hillegeist (1999)] and [Pae / Yoo (2001)] assume that the fee is yet fixed.

211
at the time the client firm chooses its action. I also make this assumption in the effort complementarity-paper. However, there are also a few models which assume that the fee is dependent on the firm’s action. A notable example is Laux / Newman (2010)\footnote{In Laux / Newman (2010) the firm’s choice variable is indeed the audit fee.}. I also make this assumption in the ICS-paper.

The attentive reader of the ICS-paper will have noticed that although the firm’s effort $s$ is observable for the risk-neutral auditor before he or she chooses $a$, the effect $\frac{\partial^2 E(U_C)}{\partial s \partial \cdot}$ is non-zero. This is because the aggregate legal liability consequences do not only consist of damages that compensate the investors for their losses, but also contain components that are real losses to the triangular auditor-client-investor economy. These costs are subsumed under the term ”litigation frictions”\footnote{See Radhakrishnan (1999) or Laux / Newman (2010).}. Typical litigation frictions include attorneys fees, legal costs, penalties, or reputation losses. Because these costs do not flow to investors, they increase the absolute liability burden of the firm, thereby making accurate reporting more important. Hence, to the extent that the legal change regards liability frictions, it alters the effort incentives of the firm, irrespective of when the fee is fixed. Notable prior papers using that mechanism are Radhakrishnan (1999), Newman / Patterson / Smith (2005), and Laux / Newman (2010). In the ICS-paper, the expected extra-liability costs towards foreseeable third parties represent such a litigation friction. By definition, the foreseeable third parties are not the immediate investors of the new investment project. Rather they are capital market participants, who sue the auditor because they did buy or sell shares of the firm on the secondary market at an unfavorable price due to their reliance on an overstated financial report. It follows that the expected liability costs from these lawsuits do not translate into a lower cost of capital for the project the firm installs the ICS for\footnote{In the overall view, also the expected extra-liability towards foreseeable third parties will affect the firm’s cost of capital (e.g. through the issuance price of new shares). However, this beneficial effect regards other financing activities not considered in the model.}. Moreover, the foreseeability approach has a substantially weaker legal foundation than the liability towards known users of financial statements. Therefore, the
extra-liability also includes the firm’s legal costs and lawyer fees associated with frivolous lawsuits.

Next, turn to the composite effect

$$\partial^2 \mathbb{E}(U^C) \partial s \partial (\cdot) = \frac{\partial^2 \mathbb{E}(U^C)}{\partial s \partial (\cdot)} + \frac{\partial^2 \mathbb{E}(U^C)}{\partial a \partial (\cdot)} \frac{\partial a^*}{\partial s} + \left[ \frac{\partial^2 \mathbb{E}(U^C)}{\partial s \partial a} + \frac{\partial^2 \mathbb{E}(U^C)}{\partial a \partial s} \right] \frac{\partial a^*}{\partial (\cdot)} + \frac{\partial \mathbb{E}(U^C)}{\partial a} \frac{\partial^2 a^*}{\partial s \partial (\cdot)}. \]

The product $\frac{\partial \mathbb{E}(U^C)}{\partial a} \frac{\partial a^*}{\partial s}$ in the firm’s FOC captures the strategic steering of the audit effort through the observable choice of $s$ in a way that optimizes the marginal contribution of $a$ to the accuracy of the report. This strategic effect can be affected by exogenous shocks like auditor liability regulation in two distinct ways. First, the change in auditor liability can directly alter the value of highly quality audits for the client firm ($\frac{\partial^2 \mathbb{E}(U^C)}{\partial a \partial (\cdot)}$). Second, it may alter the elasticity of the auditor’s effort choice to changes in firm-side effort $s$ ($\frac{\partial^2 a^*}{\partial s \partial (\cdot)}$). Of these two effects, only the former one is conditional on the design of the liability regime. The direction and the strength of the latter effect instead depends on the nature of the auditor-client collaboration, and shall be discussed later.

Again assume that $d(\cdot)$ is an increase in the damages paid to aggrieved investors in case of an audit failure. The effect $\frac{\partial^2 \mathbb{E}(U^C)}{\partial a \partial (\cdot)}$ captures whether this increase directly affects the marginal value of audit effort to the client firm. The obvious reason for a non-zero effect $\frac{\partial^2 \mathbb{E}(U^C)}{\partial a \partial (\cdot)}$ is that a part of the damages are paid by the firm. Holding the damage apportionment between the auditor and the firm constant, any increase in the overall damages increases the firm’s damage share $V_C$, and thus increases the value of diligent auditing as a measure to avoid those damages. Holding the overall damages constant, a shift in the damage apportionment towards the firm has the same effect. Now, one might argue that the effect $\frac{\partial^2 \mathbb{E}(U^C)}{\partial a \partial (\cdot)}$ should always be non-zero, because in equilibrium the firm has to compensate the auditor for his or her liability. Whether the liability change affects the firm’s expected utility directly, or through the audit fee should then not make
any difference. However, this is only true if the auditor receives a rent in equilibrium, that is the fee exceeds his or her expected disutility from the direct audit costs and the legal liability. If the audit market is perfectly price-competitive, then the contribution of the audit effort $a$ to the firm’s expected utility via the fee is already optimal, as it equals the auditor’s FOC, which is zero in equilibrium. This is the case in the ICS-paper, where the effect $\frac{\partial^2 E(U_C)}{\partial a \partial d(\cdot)}$ is non-zero solely because of the immediate effect of the change in damage apportionment on the marginal value of auditing. Conversely, if the audit market is not perfectly competitive, then the effect $\frac{\partial^2 E(U_C)}{\partial a \partial d(\cdot)}$ also contains the effect of the liability change on the portion of the fee that exceeds the auditor’s costs. This would be the case, if one implanted the supply- and demand-based pricing mechanism of Dye (1995) used in the risk aversion-paper into the sequential auditor-client models of the ICS- or the effort complementarity-paper. Finally, given that the audit market is competitive, and all the damages are paid by the auditor, the effect $\frac{\partial^2 E(U_C)}{\partial a \partial d(\cdot)}$, and thus also the composite effect $\frac{\partial^2 E(U_C)}{\partial d(\cdot) \partial a^*}$ is zero. This is the case in the effort complementarity-paper.

Similar to the just discussed effect $\frac{\partial^2 E(U_C)}{\partial a \partial d(\cdot)} \frac{\partial a^*}{\partial s}$, also the conglomerate

$$\frac{\partial^2 E(U_C)}{\partial s \partial d(\cdot)} = \frac{\partial^2 E(U_C)}{\partial d(\cdot) \partial s} + \frac{\partial^2 E(U_C)}{\partial a \partial d(\cdot)} \frac{\partial a^*}{\partial d(\cdot)} + \frac{\partial^2 E(U_C)}{\partial a \partial a^*} \frac{\partial a^*}{\partial d(\cdot)} + \frac{\partial E(U_C)}{\partial a} \frac{\partial^2 a^*}{\partial s \partial d(\cdot)}$$

regards the distribution of the liability incentives. This term depicts how the client firm’s demand for financial reporting accuracy (as given by $\frac{\partial E(U_C)}{\partial s} + \frac{\partial E(U_C)}{\partial a} \frac{\partial a^*}{\partial s}$, which is the LHS of the firm’s FOC) is affected by the direct effect of the liability change on the auditor’s effort choice ($\frac{\partial a^*}{\partial d(\cdot)}$). This effect is null if

$$\frac{\partial a^*}{\partial s} = -\frac{\frac{\partial^2 E(U_C)}{\partial s \partial a}}{\frac{\partial^2 E(U_C)}{\partial a \partial a^*}}$$

holds. Economically, this means that any effect of a change in audit effort on the firm’s
incentive to change the reporting accuracy on its own via the choice of $s$, is exactly offset by a countervailing effect on the incentive to steer the audit effort via the choice of $s$. Because in equilibrium

$$\frac{d^2 \mathbb{E}(U^A)}{da ds} = 0 \Leftrightarrow \frac{d^2 \mathbb{E}(U^A)}{\partial a \partial s} + \frac{d^2 \mathbb{E}(U^A)}{\partial^2 a} \frac{\partial a^*}{\partial s} = 0 \Leftrightarrow \frac{\partial a^*}{\partial s} = -\frac{\frac{\partial^2 \mathbb{E}(U^A)}{\partial^2 a}}{\frac{\partial^2 \mathbb{E}(U^A)}{\partial^2 a} \frac{\partial^2 \mathbb{E}(U^A)}{\partial^2 a}} \quad (5.9)$$

must hold, condition (5.8) can be rewritten as

$$\frac{\frac{\partial^2 \mathbb{E}(U^A)}{\partial a \partial s}}{\frac{\partial^2 \mathbb{E}(U^A)}{\partial^2 a}} = \frac{\frac{\partial^2 \mathbb{E}(U^C)}{\partial s \partial a}}{\frac{\partial^2 \mathbb{E}(U^C)}{\partial^2 a}}. \quad (5.10)$$

For condition (5.10) to be generally satisfied, both the numerator and the denominator of the LHS must equal their counterpart on the RHS. First, the costs and benefits that are jointly influenced by the efforts $s$ and $a$ (e.g. the expected liability payments) have to equally affect the auditor’s and the client’s expected utility (equal numerators). Second, the effect of a higher audit effort on the equilibrium cost-benefit ratio of auditing must be identical for the auditor and the client firm (equal denominators). In the present model type, the following assumptions are sufficient for the numerators and denominators of (5.10) to be equal in each equilibrium:

- The auditor’s expected costs are fully passed on to the firm via the audit fee.
- There is no effect of audit effort $a$ on the marginal costs and benefits of firm-side effort $s$, other than a change in the audit fee.
- There is no effect of audit effort $a$ on the marginal costs and benefits of audit effort $a$, other than a change in the audit fee.

These assumptions are fulfilled in the effort complementarity-paper, as well as in \textsuperscript{[Pae]} Yoo (2001). In both papers (i) only the auditor faces legal liability, (ii) the liability
is fully passed to the firm via the audit fee, and (iii) there are no spillover-effects from $a$ to any expected utility component of the firm other than the audit fee. However, in the ICS-paper the second assumption is violated, because the auditor only cares about a fraction of the total legal liability consequences, and only passes this fraction on to the firm via the fee, whereas the rest of the liability directly affects the firm. Therefore, the numerator on the RHS of condition (5.10) always has a greater absolute value than the numerator on the LHS.

5.4.2 Audit Technology

The remaining parts of the general comparative statics are the effects

$$\frac{\partial^2 E(U^C)}{ds\partial(\cdot)} = \frac{\partial^2 E(U^C)}{\partial s\partial(\cdot)} \partial a^* + \left[ \frac{\partial^2 E(U^C)}{\partial a\partial(\cdot)} \partial s + \frac{\partial^2 E(U^C)}{\partial^2 a} \partial a^* \right] \frac{\partial a^*}{\partial(\cdot)} + \frac{\partial E(U^C)}{\partial a} \frac{\partial^2 a^*}{\partial s\partial(\cdot)},$$

and

$$\frac{da^*}{d(\cdot)} = \frac{\partial a^*}{d(\cdot)} + \frac{\partial a^*}{\partial s} \frac{ds^*}{d(\cdot)}.$$

These effects regard the strategic reaction of the auditor to a change in the firm’s choice of $s$ in equilibrium. There is a common acceptance that if financial reporting is viewed as the outcome of a joint production process, firm-side investments into more accurate accounting, better internal controls, and more effective corporate governance cause a lower audit quality. The rationale is simple and compelling. If the firm does more of the work, then the auditor can afford to work less without making him- or herself worse off. In

172 In the ICS-paper I group the partial effects differently. To show the role of the strategic leverage $\frac{\partial a^*}{\partial s} < 0$ in better contour, I denote $\frac{\partial^2 E(U^C)}{\partial s\partial a} \partial a^*$ as effect (I), and summarize $\frac{\partial^2 E(U^C)}{\partial^2 a} \partial a^* \frac{\partial a^*}{\partial(\cdot)}$ and $\frac{\partial^2 E(U^C)}{\partial a\partial(\cdot)} \partial a^* \frac{\partial a^*}{\partial s}$ as effect (II).
other words, a higher firm-side effort reduces the marginal productivity of audit effort. This strategic effort-substitution, as reflected by \( \frac{2a_s}{\partial s} < 0 \), is the core of the standard multiplicative audit risk model

\[
\text{Audit Risk} = \left(1 - p\right) \ast \left(1 - s\right) \ast \left(1 - a\right) .
\]

An implicit assumption of the standard audit risk model is that the interaction between the firm’s and the auditor’s effort happens exclusively via the liability exposure. Once the auditor has assessed the firm-side risk, he or she conducts the audit without further interacting with the firm. Technically, this assumption is reflected in the multiplicative separability of the overall error risk in the auditor’s detection risk \((1 - a)\) and the firm-side risk \((1 - s)\), which implies

\[
\frac{\partial^2 (1 - d(a, s))}{\partial a \partial s} = -1.
\]

It is needless to say that in reality the auditor is not left alone after having assessed the firm-side risk. Instead, the productivity of audit effort is affected by other, more direct factors than the legal liability exposure. As an example, assume that the firm invests into a fully integrated enterprise-resource-planning (ERP) system that is capable of handling big data. This not only improves the firm’s reporting and internal control procedures, but also enhances the technology of the auditor, given that he or she gains access to these resources, and knows how to use them. Bluntly said, such an investment by the firm extends the auditor’s scope for action away from manually “rummaging” file folders on a random sample basis, towards a (semi-automated) complete audit. While only few auditors have clients with state-of-the-art data superhighways, the effect is qualitatively the same if the firm’s interaction with the auditor is limited to occasional testing assistance or the provision of carefully prepared books. Firm-side investments in financial accounting
and reporting increase the amount of assurance the auditor can gain by auditing one more hour. The positive effect of effort $s$ on the marginal productivity of audit effort $a$ is called \textit{strategic effort complementarity} and is formally reflected by $\frac{\partial a^*}{\partial s} > 0$.

Whether the overall effect of firm-side accounting and control effort on the marginal productivity of auditing is positive ($\frac{\partial a^*}{\partial s} > 0$) or negative ($\frac{\partial a^*}{\partial s} < 0$), that is whether the two efforts are strategic complements or substitutes in equilibrium, depends on how strong the complementary technology-spillover is relative to the absolute reduction of the error risk by the firm. This is ultimately an empirical question. To give comprehensive guidance for empirical observations and statistical inferences, theory must at least allow for a complementary effect of firm-side effort on the productivity of audit effort. However, by assuming a multiplicative linkage between the firm’s and the auditor’s efforts, this effect is categorically excluded throughout the literature.\footnote{Examples include Nelson / Ronen / White (1988), Smith / Tiras / Vichitlekarn (2000), and Pae / Yoo (2001).} This assumption creates an incomplete picture of auditor-client interaction. It is thus unacceptable when it comes to drawing theoretical predictions for empirical research on the relative resource allocation between the auditor and the firm in the reporting process.\footnote{Still the assumption of exclusive strategic substitution is as acceptable as most other simplifying assumptions when the goal is to work out selected incentive effects as crisp as possible.}

From a modeling perspective, incorporating the aspect of strategic effort complementarity requires that the change of the slope coefficient of the audit technology $\frac{\partial d(a,s)}{\partial a}$ in firm-side effort $s$, that is the partial derivative $\frac{\partial^2 d(a,s)}{\partial a \partial s}$, has both a positive and a negative component. As shown above, this is not true for the multiplicative audit technology $d(a,s) = 1 - (1 - s)(1 - a)$. In the effort complementarity-paper, I use the negative exponential audit technology $d(a,s) = 1 - e^{-sa}$. One can readily see from

$$\frac{\partial^2 (1 - d(a,s))}{\partial a \partial s} = e^{-sa}(1 - sa),$$

218
that this technology fulfills the requirement of having both a negative and a positive component. The same holds true for the square root technology

\[ d(a, s) = \sqrt{sa}, \]

and various logarithmic specifications, such as

\[ d(a, s) = s \ast \ln(a), \]

or

\[ d(a, s) = \ln(s) \ast \ln(a). \]

While all these functions basically allow for both strategic complementarity and substitution, only the exponential specification actually exhibits both strategic relation types in equilibrium. For sufficiently low levels of \( s \), the strategic complementarity effect dominates, whereas for higher levels of \( s \) the substitution effect prevails. The short empirical cross-country analysis in the effort complementarity-paper suggests that this progression is descriptive of reality. Under the square root and logarithmic specifications the complementarity effect always dominates, which means that in every equilibrium firm-side effort \( s \) and audit effort \( a \) are strategic complements. Such a unidirectional dominance is generally unrealistic. However, there might still be use for those functions, as will become clear in a moment.

The strategic effect \( \frac{\partial a^*}{\partial s} \) is critical, because it determines the direction of the auditor’s reaction to the client firm’s choice of \( s \). For the firm, the effect \( \frac{\partial a^*}{\partial s} \) is less distinctive, since it is only a part of the second-order effect \( \left[ \frac{\partial^2 \mathcal{E}(U^C)}{\partial a \partial a} + \frac{\partial^2 \mathcal{E}(U^C)}{\partial a^2} \frac{\partial a^*}{\partial s} \right] \frac{\partial a^*}{\partial \mathcal{U}}. \) This effect is in turn only one of three second-order effects on the firm’s decision problem, and under
certain conditions is even null (see previous subsection). Because the firm is the first-mover (Stackelberg-leader) of the game, it already optimizes over the strategic effect \( \frac{\partial a^*}{\partial s} \). However, what the firm cares about is in which direction and to which extent the change in the liability regime affects the strategic effect, and thus requires a re-optimization over \( \frac{\partial a^*}{\partial s} \). This incentive is captured by the product \( \frac{\partial \mathbb{E}(U_C)}{\partial a} \frac{\partial^2 a^*}{\partial s(\cdot)} \).

The partial derivative \( \frac{\partial \mathbb{E}(U_C)}{\partial a} \) captures whether external auditing is a conventional complement to the firm’s effort \( s \). The two efforts are conventional complements if \( a \) has a positive effect on the total productivity of \( s \) et vice versa. This is true in all models in which the firm and the auditor are both interested in producing accurate financial statements. Conversely, if the firm plays “against” the auditor by consciously biasing the report, the firm’s manipulation effort and the auditor’s detection effort are conventional substitutes. While I explicitly focus on a scenario in which the firm and the auditor pull in the same direction, the model (5.1)-(5.4) is also capable of describing the interaction between the auditor and an intentionally misreporting client. Whether the firm consciously overstates the report to get some benefit (such as in Hilgegeist (1999)), or refrains from detecting existent overstatements in order to save control costs (such as in Pae / Yoo (2001) or the ICS-paper), basically makes no difference. In both cases the accuracy of published financial reports is the result of a tradeoff between the costs and benefits of overstatements.

The effect \( \frac{\partial^2 a^*}{\partial s(\cdot)} \) depicts the elasticity of the strategic effect \( \frac{\partial a^*}{\partial s} \) w.r.t. changes in exogenous parameters. Economically, the effect is difficult to grasp.\(^{175}\) The trigger of the effect is a first-order effect of an exogenous shock on the equilibrium audit effort \( a^* \) (e.g. higher auditor liability). Sticking to our example, let the first-order effect of liability consequences on the equilibrium audit effort be positive, that is a higher liability threat makes the

\(^{175}\) For example, Pae / Yoo (2001) introduces this effect as "the impact of the auditor’s expected liability loss, author’s note on the slope of the auditor’s best response function" (Pae / Yoo (2001), p 345). They then impose \( \frac{\partial^2 a^*}{\partial s(\cdot)} \leq 0 \) as a ”regularity condition” without providing an economic justification.
The sign of $\frac{\partial^2 a^*}{\partial s \partial V}$ now depends on whether the increase in audit effort makes the collaboration more or less cost-efficient. If the higher audit effort makes the reliance on firm-side activities $s$ less valuable or even causes testing redundancies, then $\frac{\partial^2 a^*}{\partial s \partial V}$ is negative, and the firm has an incentive to reduce $s$. However, if the auditor can utilize the firm’s preliminary work better if he or she works harder, then $\frac{\partial^2 a^*}{\partial s \partial V}$ is positive. The following table summarizes the properties of the discussed audit technologies:

<table>
<thead>
<tr>
<th>$d(a, s)$</th>
<th>$\frac{\partial a^*}{\partial s}$</th>
<th>$\frac{\partial^2 a^*}{\partial s \partial V}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1 - (1 - s)(1 - a)$</td>
<td>$&lt; 0$</td>
<td>$&lt; 0$</td>
</tr>
<tr>
<td>$1 - e^{-sa}$</td>
<td>$&gt; 0/ &lt; 0$</td>
<td>$&lt; 0$</td>
</tr>
<tr>
<td>$\sqrt{sa}$</td>
<td>$&gt; 0$</td>
<td>$&gt; 0$</td>
</tr>
<tr>
<td>$s * ln(a)$</td>
<td>$&gt; 0$</td>
<td>$&gt; 0$</td>
</tr>
<tr>
<td>$ln(s) * ln(a)$</td>
<td>$&gt; 0$</td>
<td>$&gt; 0$</td>
</tr>
</tbody>
</table>

In reality, the effect $\frac{\partial^2 a^*}{\partial s \partial V}$ will be both positive and negative. If the auditor exerts very little audit effort, it is unlikely that he or she can utilize the resources provided by the firm’s investment $s$. This would require a minimum amount of audit effort. Therefore, the effect $\frac{\partial a^*}{\partial s}$ is likely to be increasing in $a^*$, respectively in any exogenous parameter that increases $a^*$, for sufficiently low $a^*$. However, as the supplied audit quality increases, the efficiency gains are driven out by the emergence of testing- and control-redundancies. This means that the positive effect $\frac{\partial a^*}{\partial s}$ fades out, and ultimately turns negative for sufficiently high $a^*$. While the negative exponential technology $d(a, s) = 1 - e^{-sa}$ captures the latter transition from strategic complementarity to strategic substitution well, it does not account for the

---

176 The risk aversion-paper shows that this is not necessarily always the case.

177 Technically, the effect is positive if the change of the slope coefficient $\frac{\partial d(a, s)}{\partial a}$ in $s$ becomes more positive (less negative) as $a$ increases, given that the exogenous factor $(\cdot)$ increases $a$. Graphically, a positive (negative) effect can be recognized by the spreading-out (convergence) of the iso-curves of $d(a, s)$ for different levels of $s$, in $a$. For example, the iso-curves of the negative exponential audit technology $d(a, s) = 1 - e^{-sa}$ converge as $a$ increases (see figure 4.2 in the effort complementarity-paper), which is indicative of $\frac{\partial^2 a^*}{\partial s \partial V} < 0$, given that $\frac{\partial a^*}{\partial s}$ is positive. Conversely, the iso-curves of the square root audit technology $d(a, s) = \sqrt{sa}$ spread out as $a$ increases, which is indicative of $\frac{\partial^2 a^*}{\partial s \partial V} > 0$ (see figure 4.3 in the effort complementarity-paper).
CONCLUSION

initial emergence, respectively reinforcement of the strategic complementarity through improvements in poor quality audits. Instead, it assumes the strategic complementarity to be strongest for the worst quality audits. Conversely, the square root specification $d(a, s) = \sqrt{sa}$ or the logarithmic specifications mentioned above capture the reinforcement of strategic complementarity through improvements in poor quality audits, but totally fail to account for the ultimately prevailing strategic substitution.

The obvious way to implement the hump-shaped progression of $\frac{\partial^2 a^*}{\partial s \partial V}$ into the model is to modify the negative exponential technology, which already captures the transition from strategic complementarity to strategic substitution, by adding a power component, which brings along the desired property $\frac{\partial^2 a^*}{\partial s \partial V} > 0$. For example, consider the function class

$$d(a, s) = 1 - e^{-\alpha s \beta a} + \sqrt{a},$$

with $\alpha$ and $\beta$ being positive constants. This function is strictly increasing and non-convex in both $a$ and $s$. Moreover, it can be easily verified by using condition (5.9), that given an appropriate parametrization, $a$ and $s$ are strategic complements for low levels of $s$, and strategic substitutes for higher levels of $s$. Finally, by differentiating the RHS of (5.9) w.r.t $a$, one can see that the function is increasing in $a^*$ for low levels of $a^*$, but decreasing in $a^*$ for higher levels of $a^*$. Because $a^*$ is strictly increasing in $V$ by assumption, this means that $\frac{\partial^2 a^*}{\partial s \partial V}$ is positive for sufficiently small $a^*$, but negative for sufficiently large $a^*$. The model described by (5.1)-(5.4) has well enough degrees of freedom, such that a set of parameters which gives these progressions of the effect $\frac{\partial a^*}{\partial s}$ for the proposed modification always exists. On the downside, the modification prohibits the explicit solution for $a^*$

---

178 This is true, because ceteris paribus any level of $s^*$ or $a^*$ can be implemented by varying the cost functions $c(s)$ and $k(a)$.
at the second stage of the game.\footnote{To get a simple explicit solution with the negative exponential specification, \( a \) must only appear in the exponent \( e^{-\alpha s \beta a} \) of the expected liability term, whereas \( \alpha \) and \( \beta \) are positive constants. Otherwise, the explicit analytical solution of \( a^* \) includes a Lambert-W function, which is difficult to use in the comparative statics. However, given that \( a \) appears solely in the exponent, the effect \( \frac{\partial a^*}{\partial s} \) has the form \( \frac{1}{\ln[f(a,s)]} \) with \( f(a,s) \) and \( g(s) \) being both strictly increasing in \( a \) and \( s \). It follows that any exogenous parameter that has a monotone effect on \( a^* \), also has a monotone effect on \( \frac{\partial a^*}{\partial s} \).} Having only an implicit second-stage solution makes it very hard to analytically quantify the strength of countervailing firm incentives at the first stage of the game. For general comparative static analyses in sequential, or even multi-period settings, it is thus preferable to work with a piecewise audit technology, such as

\[
d(a, s) = \begin{cases} \sqrt{sa} & \text{if } s \leq \hat{s} \\ 1 - e^{-sa} & \text{if } s > \hat{s}, \end{cases}
\]

with \( \hat{s} \) being some threshold value. A piecewise modeling is also preferable for studying the strategic auditor-client-interaction in two "extreme" institutional environments, i.e. completely different levels of \( s \). Conversely, if it is ex ante unclear which case regarding \( \frac{\partial a^*}{\partial s} \) and \( \frac{\partial^2 a^*}{\partial s \partial (\cdot)} \) prevails, it is recommendable to parametrize the model as much as possible, and work with an "all-in-one" modeling such as (5.11). This particularly applies to empirical studies.

### 5.5 Concluding Remarks

Having disassembled the standard sequential auditor-client model to a considerable level of detail, it is obvious that there is no clear answer to the main research question outlined in the introductory chapter. This is somehow unsatisfactory, but no surprise at all. However, what may be a surprise is to see how many different factors influence the aggregate effect of legal liability regulation on the strategic auditor-client-interaction even in a simple, stylized model economy. The following list summarizes these factors in terms of questions to ask when approaching this topic with an economic model:
CONCLUSION

• Is the auditor and/or the firm risk-neutral, or risk-averse?

• Does the firm consider any costs and benefits from auditing that the auditor does not consider?

• Is the audit market perfectly price-competitive, or do auditors earn rents?

• Is the audit fee fixed before or after the auditor observes the firm’s action?

• Are the firm’s accounting-related efforts, and the auditor’s effort conventional substitutes or conventional complements?

• Are the firm’s accounting-related efforts, and the auditor’s effort strategic substitutes or strategic complements?

• Is the elasticity of the strategic substitution/complementarity between the firm’s and the auditor’s efforts w.r.t. audit effort positive, or negative?

I conclude by reflecting on the implications of this dissertation for research and practice. The main goal of any research in the field of accounting and auditing is to give serious advice to regulators and policymakers. Each of the dissertation papers is set up with this ambition. However, regulators should be careful when using this dissertation (and any other piece of theoretical accounting research) as an immediate basis for action. First, none of the papers makes a statement about the importance of the proposed effects in reality. While the existence of all the new and unexpected effects in the papers is a logical consequence of reasonable model assumptions, it is unclear whether they are economically significant. Second, the direction of most effects hinges on the economic environment, as represented by certain parameter constellations. None of the models makes a definitive statement about which parameter constellation describes which economy. Answering these

180 I leave it to the reader to judge whether this is an immediate threat.
two questions, and thereby judging whether the results should be considered in regulation or not, requires explicit empirical diagnostics.

For empirical research, this dissertation offers numerous starting points. Most obviously, it contains an array of novel theoretical predictions about yet unexplored, counter-intuitive economic effects. With access to state-of-the-art financial statement- and auditing databases, most of these predictions can be directly tested. However, the dissertation does not only serve as a source for new empirical research. More importantly, it provides a comprehensive theory of the sequential auditor-client collaboration in the financial reporting process. This theory as a whole should help us to better understand existent empirical evidence about auditor-client interaction. The review of the empirical literature in the effort complementarity-paper suggests that a solid theoretical foundation of this topic is urgently needed, as most of the statistical analyses out there seem to be based on intuitive conjectures. Given that ten different partial effects (summarized in the equations (5.5) and (5.7)) have been identified as determinants of the auditor’s and the client’s strategic behavior, the “intuition-only-approach” in the interpretation of statistical associations is likely to have the accuracy of a shotgun. Especially the firm’s first-stage decision is the result of an interplay between several partial effects, which are fairly subtle, and thus easily disregarded. Moreover, the majority of the effects does not have an intuitive direction at all. As all these effects flow together in the auditor’s second-stage decision, an intuitive first-order effect (such as the positive impact of a the firm’s damage share on internal control quality) can quickly get lost in the flood of second-order effects. In addition, there can be counter-intuitive first-order effects, such as the ones discussed in the risk aversion-paper. Finally, it might even be that two counter-intuitive effects offset each other, such that the observable outcome appears intuitive again. For example, an overall negative effect of a damage shift from the auditor towards the firm on

\[ \text{It is however difficult to test the results regarding the impact of regulation on social welfare, due to the lack of a distinct proxy for social welfare.} \]
internal control effort has a negative effect on the supplied audit quality under strategic complementarity. However, so does a positive effect of the damage shift under strategic substitution. Without having a clear idea about the incentive structure in the sample and the population, one cannot say which of the arguments explains an observable reduction of audit quality.

I am aware that grasping subtle second-order effects with econometric or experimental instruments is a tedious task, which few empiricists will pick up. I am also aware that the direct discernible impact of the three dissertation papers on regulation will probably be low, not to say close to zero. However, I believe that this dissertation is a significant addition to the ever-growing body of theoretical accounting and auditing literature, and as such will have an impact on research and practice. Apart from presenting several novel, and in my view intriguing results, it structures and integrates existent theoretical pieces of research in the field. This sharpens the awareness for potentially fruitful future research topics. In particular, three fields seem promising to me: First, and quite obvious, the enforcement system, and in particular the distinct roles of different enforcement institutions are to be modeled more comprehensively. For example, the recent paper by Ewert / Wagenhofer (2018) studies the joint effect of public enforcement and auditing on financial reporting accuracy, taking into account the differences in the tasks of an auditor and an enforcer. Second, more theoretical research on the structure of the audit market, and the evolution of the prevalent market concentration is warranted. This is a red-hot topic among regulators, which is clearly under-researched. The framework of Dye (1993) and Dye (1995) is the only comprehensive market model for more than twenty years now, and yet it has been barely picked up. Beyond that, I am convinced that the insights, and more importantly, the impact of audit-market research, exponentially increase from the point of time when the research community taps into the general economic theory of markets and competition on a large scale. Third and finally, I think that the real effects of accounting and auditing on the markets for goods, services and labor should be taken
CONCLUSION

into account much more. In a current working paper (Hinterecker / Schrank (2019)) we study how tightening a liability regime affects a firm’s investment policy, production quantities, as well as the producer and consumer surplus under different types of quantity competition. This broad view allows us to make statements about the "macro-effects" of regulation, and builds a bridge between auditing and industrial economics research. In the same spirit, problems of accounting and auditing can be connected to problems of operations research, human resource management, or marketing. Not only is this kind of research likely to yield tons of intriguing new insights that will find their way into regulation and practice. Above all, this kind of research will confidently position the field of financial accounting, reporting and auditing where it is in the corporate world, and where it should also be in business research - at the very heart.

The working paper is available upon request.
References


Bigus, J. (2011): Auditors’ liability with overcompensation and reputation losses,


REFERENCES


XIII


Smith, J./ Tiras, S./ Vichitlekarn, S. (2000): The Interaction between Internal Con-
REFERENCES


