

The oversight of financial reporting in Germany:

Essays on audit quality and on accounting enforcement

Dissertation

zur Erlangung des Doktorgrades

der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Camilla Fiallo

Tübingen

2019

Tag der mündlichen Prüfung:

12.06.2020

Dekan:

Prof. Dr. Josef Schmid

1. Gutachter:

Prof. Dr. Renate Hecker

2. Gutachter:

Prof. Dr. Werner Neus

Index

INDEX	I
LIST OF FIGURES.....	V
LIST OF TABLES	VI
LIST OF ABBREVIATIONS	VIII
1. INTRODUCTION.....	1
2. AUSGEWÄHLTE NEUREGELUNGEN DER EU-VERORDNUNG ZUR ABSCHLUSSPRÜFUNG UND PRÜFUNGSQUALITÄT (TEIL 1, 2 UND 3) – EINE EMPIRISCHE BESTANDSAUFNAHME BEI DEUTSCHEN UNTERNEHMEN MIT DPR- FEHLERFESTSTELLUNGEN	6
2.1 Einleitung	7
2.2 Fehlerveröffentlichungen nach Abschluss eines Enforcementverfahrens als Surrogatmaß für Prüfungsqualität	8
2.3 Datenbasis und Abgrenzung der Stichprobe	13
2.4 Untersuchungsergebnisse – Ist die Verschärfung des Verbots von Nichtprüfungsleistungen unter dem Gesichtspunkt der Prüfungsqualität sinnvoll?.....	18
2.5 Zusammenfassung zu Teil 1 und Ausblick	28
2.6 Einleitung zu Teil 2	29
2.7 Untersuchungsergebnisse – Obergrenzen für die Honorare aus Nichtprüfungsleistungen und für die Gesamthonorare.....	31
2.7.1 Ist die Einführung des cap eine wirksame Maßnahme zur Steigerung der Prüfungsqualität?.....	31
2.7.2 Ist die 15%-Honorargrenze der EU-VO effektiv im Hinblick auf die Steigerung der Prüfungsqualität?.....	38
2.8 Zusammenfassung zu Teil 2 und Ausblick	46
2.9 Einleitung zu Teil 3	47
2.10 Untersuchungsergebnisse – Externe Rotation.....	49
2.11 Zusammenfassung und Einschränkungen der Untersuchung.....	56

3. THE PROVISION OF NON-AUDIT SERVICES UNDER THE EU-REGULATION ON THE STATUTORY AUDIT OF PIES: EVIDENCE BASED ON ENFORCEMENT ANNOUNCEMENTS IN GERMANY	61
3.1 Introduction	62
3.2 Institutional framework	65
3.3 Theoretical background, literature review and hypothesis development	67
3.3.1 Theoretical background	67
3.3.1.1 Audit quality and auditor independence	67
3.3.1.2 Economic bonding and knowledge spillovers	68
3.3.2 Prior literature	71
3.3.2.1 The provision of NAS and independence in appearance	71
3.3.2.2 The provision of NAS and independence of mind	72
3.3.2.3 Future fees and audit quality	74
3.3.2.4 NAS categories and audit quality	74
3.3.2.5 Client importance	76
3.3.3 Research hypotheses	78
3.4 Research design	84
3.4.1 Surrogates for audit quality	84
3.4.2 Sample selection	88
3.4.3 Model and main variables	91
3.4.4 Control variables	96
3.5 Results	99
3.5.1 Descriptive statistics	99
3.5.2 Multivariate results	103
3.5.2.1 H1: NAS categories	103
3.5.2.2 H2: the cap	109
3.5.2.3 H3: Client importance	111
3.5.3 Sensitivity tests	113
3.5.3.1 Financial companies	113
3.5.3.2 Endogeneity	115
3.5.3.3 Size	116
3.6 Conclusion	117

4. FUTURE FEES, CURRENT NAS FEES AND AUDIT QUALITY: EVIDENCE BASED ON ENFORCEMENT ANNOUNCEMENTS IN GERMANY	121
4.1 Introduction	122
4.2 Theoretical background, research hypotheses and prior research	124
4.2.1 Theoretical background and hypotheses	124
4.2.2 Prior literature	127
4.3 Research design	129
4.3.1 Sample selection	129
4.3.2 The model	130
4.4 Results	132
4.4.1 Descriptive statistics	132
4.4.2 Multivariate results	135
4.4.3 Sensitivity tests	138
4.4.3.1 Financial companies	138
4.4.3.2 Auditor change	138
4.4.3.3 Client importance	141
4.4.3.4 Time frame	142
4.5 Conclusion	145
5. EARNINGS MANAGEMENT AND THE GERMAN ENFORCEMENT SYSTEM: AN APPLICATION OF THE M-SCORE	148
5.1 Introduction	149
5.2 Institutional framework	152
5.2.1 The German enforcement system	152
5.2.2 The work of the FREP	155
5.3 Theoretical background and hypotheses	157
5.3.1 Theoretical background	157
5.3.2 Research hypotheses	158
5.3.2.1 Enforcement announcements and earnings management	158
5.3.2.2 Error severity	160
5.3.2.3 Cooperation of the examined company	162
5.4 The model of Beneish (1999)	165

5.4.1	Measuring earnings management.....	165
5.4.2	The m-score.....	165
5.4.3	Prior studies using the m-score	169
5.4.4	Brief discussion on the need to re-estimate the model of Beneish (1999).....	171
5.5	Research design.....	173
5.5.1	Sample selection.....	173
5.5.2	Methodology	176
5.6	Results	178
5.6.1	Univariate results: enforcement announcements and earnings management.....	178
5.6.2	Univariate results: error severity and cooperation of the examined company	182
5.6.3	Multivariate results.....	187
5.7	Conclusion.....	190
6.	SUMMARY AND CONCLUSION.....	192
	REFERENCES.....	195

List of Figures

Abb. 2.1: Paarweiser Vergleich der Einhaltung der 15%-Obergrenze im Jahr des Fehlers	44
Figure 3.1: Definition of audit failure	85
Figure 4.1: Audit quality and the provision of NAS in a multi-period framework.....	126

List of Tables

Tab. 2.1: Ergebnisse der Abschlussprüfung der fehlerhaften Abschlüsse und deren Verteilung über die Zeit.....	11
Tab. 2.2: Abgrenzung der Stichprobe	13
Tab. 2.3: Häufigkeit des Auftretens von Abschlussprüfungs- und Nichtprüfungsleistungen im Jahr des Fehlers.....	24
Tab. 2.4: Paarweiser Vergleich von Abschlussprüfungs- und Nichtprüfungsleistungen im Jahr des Fehlers	27
Tab. 2.5: Überschreitung des cap im Jahr des Fehlers	34
Tab. 2.6: Paarweiser Vergleich der Einhaltung des cap im Jahr des Fehlers ^{a)}	37
Tab. 2.7: Überschreitung der 15%-Honorargrenze im Jahr des Fehlers	42
Tab. 2.8: Prüferwechsel im Jahr des Fehlers oder in den zwei Jahren davor.....	53
Tab. 2.9: Paarweiser Vergleich von Prüferwechseln im Jahr des Fehlers oder in den zwei Jahren davor.....	54
Table 3.1: Sample selection	89
Table 3.2: Sector and year distribution	90
Table 3.3: Sample selection - H2	94
Table 3.4: Sample selection - H3	95
Table 3.5: Variables	98
Table 3.6: Descriptive statistics	100
Table 3.7: Pearson's and Spearman's correlation coefficients (part 1).....	101
Table 3.8: Pearson's and Spearman's correlation coefficients (part 2).....	102
Table 3.9: Multivariate results for H1- part 1	106
Table 3.10: Multivariate results for H1 - part 2	107
Table 3.11: Multivariate results for H1 - with interaction term	108
Table 3.12: Multivariate results for H2.....	110
Table 3.13: Multivariate results for H3.....	112
Table 3.14: Multivariate results for H1 - without financial companies – part 1	114
Table 3.15: Multivariate results for H1 - without financial companies – part 2	114
Table 3.16: ITCV for M1 (NAScat_c, NAScat_d, NASratio_c, AuditFees).....	116
Table 4.1: Sample selection	129

Table 4.2: Sector and year distribution	130
Table 4.3: Fee variables	131
Table 4.4: Control variables	132
Table 4.5: Descriptive statistics	133
Table 4.6: Pearson's and Spearman's correlation coefficients (part 1).....	134
Table 4.7: Pearson's and Spearman's correlation coefficients (part 2).....	135
Table 4.8: Multivariate results	137
Table 4.9: Multivariate results - without financial companies.....	139
Table 4.10: Multivariate results - without subsequent auditor changes and with client importance	140
Table 4.11: Multivariate results - one-year period.....	143
Table 4.12: Multivariate results - three-year period.....	144
Table 5.1: Overview of the BaFin examinations from July 2005 until December 2017	163
Table 5.2: Variables	168
Table 5.3: Sample selection - Error group	174
Table 5.4: Sample selection - Non-error group.....	175
Table 5.5: Sector distribution	175
Table 5.6: Year distribution	176
Table 5.7: Comparison in terms of size, profitability and leverage in the error year	176
Table 5.8: Classification according to the m-score per year and sector - H1	180
Table 5.9: Descriptive statistics - H1	181
Table 5.10: Univariate regressions - H1.....	182
Table 5.11: Classification according to the m-score per year and sector - H2	183
Table 5.12: Classification according to the m-score per year and sector - H3	184
Table 5.13: Descriptive statistics - H2 and H3.....	185
Table 5.14: Univariate regressions - H2.....	186
Table 5.15: Univariate regressions - H3.....	187
Table 5.16: Multivariate regressions with and without year and sector fixed effects - H1	188
Table 5.17: Multivariate regressions without year and sector fixed effects - H2 and H3.....	189

List of Abbreviations

a.a.O.	am angegebenen Ort
ABIEU	Amtsblatt der Europäischen Union
aBL	andere Bestätigungsleistungen
AktG	Aktiengesetz
AOB	Auditor Oversight Body
APAS	Abschlussprüferaufsichtsstelle
APL	Abschlussprüfungsleistungen
AReG	Abschlussprüfungsreformgesetz
BaFin	Bundesanstalt für Finanzdienstleistungsaufsicht
BilMoG	Bilanzrechtsmodernisierungsgesetz
CESR	the Committee of European Securities Regulators
DPR	Deutsche Prüfstelle für Rechnungslegung
EGHGB	Einführungsgesetz zum Handelsgesetzbuch
ESMA	European Securities and Markets Authority
EU	European Union
EU-Reg.	Regulation (EU) No 537/2014
EU-RL	Richtlinie 2014/56/EU
EU-VO	Verordnung (EU) Nr. 537/2014
FREP	Financial Reporting Enforcement Panel
GAAP	generally accepted accounting principles
GCO	going concern opinion
GICS	Global Industry Classification Standard
HGB	Handelsgesetzbuch
IAS	International Accounting Standard
IDW	Institut der Wirtschaftsprüfer
IDW PS	IDW Prüfungsstandard
IDW RS HFA	IDW Stellungnahme zur Rechnungslegung Hauptfachausschuss
IFRS	International Financial Reporting Standards
ISA	International Standards on Auditing
ITCV	impact threshold for a confounding variable

KapGes	Kapitalgesellschaft
KoR	Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung
NAS	non-audit services
NPL	Nichtprüfungsleistungen
OR	odds ratio
PP&E	property, plant and equipment
PW	Prüferwechsel
RoA	return on assets
RoE	return on equity
SEC	Securities and Exchange Commission
sL	sonstige Leistungen
SOX	Sarbanes Oxley Act
StBL	Steuerberatungsleistungen
VIF	Variance Inflation Factor
WESML	weighted exogenous sample maximum likelihood probit
WpHG	Wertpapierhandelsgesetz
WPO	Wirtschaftsprüferordnung

1. Introduction

The oversight of financial reporting plays a fundamental role in the stability of the financial market. Financial statements are the primary communication tools between the company's representatives and the company's stakeholders and are the basis for economic decisions, which is why it is of utmost importance that the reporting is accurate and reliable. Generally, financial statements of publicly traded companies are prepared by the management under the oversight of the supervisory board, and are subject to a statutory audit, which attests whether the financial statements are a fair presentation of the company's financial situation and whether they have been prepared in accordance with the applicable accounting standards¹. In the last two decades, some major accounting scandals as well as the financial crisis have shown the need of further improving the integrity and reliability of financial reporting². In response to this, several regulative actions have been undertaken, both at European and at national level.

To enhance the quality of financial statements, achieve financial stability and restore market confidence, the European Union (EU) has promoted the institution of national enforcement systems, which oversee the compliance of financial information with the applicable reporting framework³. The two-tier German enforcement system was introduced in 2004 under the Financial Reporting Compliance Act (Bilanzkontrollgesetz). The Financial Reporting Enforcement Panel (FREP) and the German Federal Financial Supervisory Authority (BaFin) monitor the reporting of financial information, by examining financial statements and the respective management reports of German companies, which have securities admitted to trading in a regulated segment of the national market⁴. To increase the transparency of financial information, material accounting errors found during an enforcement examination must be disclosed in an enforcement announcement.

To enhance the quality of financial statements and strengthen the role of the statutory auditor, the EU issued, in April 2014, Directive 2014/56/EU, which amends the existing Directive 2006/43/EC,

¹ ISA 200.3

² E.g. Marten/Quick/Ruhnke (2015), p. 2.

³ The Committee of European Securities Regulators (CESR), Standard no. 1 on financial information - enforcement of standards on financial information in Europe, pp. 3-4. In 2011 the CESR was replaced by the European Securities and Markets Authority (ESMA).

⁴ Sec. 342b para. 2 HGB. For a more detailed analysis of the German enforcement system, see section 5.2.

and Regulation (EU) No 537/2014⁵. The EU-regulation, which applies to the statutory audit of public interest entities (PIEs⁶), aims at improving auditor independence and audit quality by restricting the provision of certain non-audit services (NAS), by setting a cap on NAS fees as well as on the proportion of fees an auditor can obtain from a client, and by limiting the length of audit engagements⁷. Whether these measures are useful in reducing the risk of economic and social bonding and in improving auditor independence is controversial.

The risk of economic bonding is explained by the quasi-rent theory of *DeAngelo* (1981a)⁸. While the initial audit might be characterised by fees below the actual cost of audit (low balling), subsequent audits generate positive quasi-rents, due to the cost advantage the incumbent auditor has compared to other auditors. It follows that the incumbent auditor is interested in keeping the client, which might threaten his independence. The provision of NAS might increase the value of the quasi-rents, as the auditor has an additional cost advantage, thus strengthening the economic bond between auditor and client⁹. However, the additional knowledge generated by NAS (so called knowledge spillovers) might boost the ability of the auditor to detect errors¹⁰. It follows that, at least in theory, the provision of NAS might both improve and worsen audit quality.

The risk of social bonding is related to the auditor developing a close relationship with the client, which might reduce the auditor's objectivity and professional scepticism. The provision of NAS might contribute to social bonding¹¹. In fact, the advisory activity frequently involves an intensive collaboration between the auditor and the company's representatives (familiarity threat), while the provision of services that involve promoting or defending a client's position might lead the auditor to identifying with the client (advocacy threat)¹². Typically, the familiarity threat arises when the

⁵ Hereafter called EU-regulation or EU-reg.

⁶ The term PIE refers to capital market-oriented companies that have already been admitted to trading in an EU state, as well as certain banks and insurance companies. For a more detailed definition of PIE as well as references to the regulatory framework, see e.g. Fiallo/Hecker (2019a), p. 225 footnote 2.

⁷ Recitals 7, 8, 21 EU-Reg.

⁸ DeAngelo (1981a), pp. 113-127. This theory, as well as its extensions, are discussed in more detail in section 3.3.1.2.

⁹ Ostrowski/Söder (1999), p. 562. On the interaction between auditor independence (or lack thereof) and audit quality, see also Fiallo/Hecker (2019a), p. 226.

¹⁰ Simunic (1984), pp. 680-681. On the potential benefits of knowledge spillovers on audit quality, see Fiallo/Hecker (2019a), p. 231. A more detailed analysis of the theoretical arguments beyond knowledge spillovers is included in section 3.3.1.2 of this study.

¹¹ E.g. Hohenfels/Quick (2018), OnlineFirst 25 October 2018. Available at:

<https://link.springer.com/article/10.1007/s11846-018-0306-z>.

¹² The five threats to independence are listed in IESBA (2018), sec. 120.6 A3, and are discussed in section 3.3.1.1.

auditor and the client work together for many years, which is an argument in favour of limiting the duration of audit engagements¹³. However, the opposing argument reads that audit quality might decrease as the new auditor does not have previous knowledge or experience with the client¹⁴.

In this dissertation I analyse several aspects related to the oversight of financial information in Germany. In chapter 2, I examine, together with Renate Hecker, whether certain limitations introduced by the EU-regulation are potentially able to improve audit quality. Audit quality is measured with the existence in audited financial statements of accounting errors, which are identified thanks to the enforcement announcements disclosed in the Federal Gazette. We run a series of descriptive and univariate tests trying to shed some light on the following issues. First, the EU regulator has expressed concerns that the provision of certain NAS might compromise auditor independence¹⁵, therefore we analyse if and how certain types of services might affect audit quality. Second, as the level of fees and their structure are suspected of impairing auditor independence¹⁶, we analyse whether violating the cap on NAS fees introduced by the EU-regulation is associated with lower audit quality. Third, we address the threat identified by the EU regulator of an auditor becoming excessively dependent on a client¹⁷, by testing if the auditor is more likely to impair his independence when he obtains more than 15% of his revenue from a client. Fourth, while limiting the duration of the audit engagement might reduce the risk of social bonding¹⁸, audit quality might decrease as the new auditor does not have previous knowledge or experience with the client. Therefore, we analyse whether there is evidence of lower audit quality in the first engagement years.

In chapter 3, I build on and complement the descriptive and univariate analyses carried out in chapter 2 by focusing on the risk of economic bonding. I further examine whether limiting the provision of certain NAS and setting a cap on NAS fees as well as on the proportion of fees an auditor can obtain from a client might improve audit quality. After reviewing prior studies and

¹³ Recital 21 EU-Reg. Limiting the duration of the engagement also reduces the present value of the quasi-rents. Whether or not this lowers the risk of economic bonding is not clear, as due to the mandatory rotation, the present value of the quasi-rents is reduced also for other engagements. See here the discussion in Fiallo/Hecker (2019c), p. 344 as well as in Wagenhofer/Ewert (2015), p. 529-530.

¹⁴ E.g. Fiallo/Hecker (2019c), p. 344; Quick and Wiemann (2011), p. 934; Krauß/Zülch (2013), p. 316; Marten/Quick/Ruhnke (2015), p. 299.

¹⁵ Recital 7 EU-Reg.

¹⁶ Recital 7 EU-Reg.

¹⁷ Recital 7 EU-Reg.

¹⁸ Recital 21 EU-Reg.

describing the quasi-rents and the knowledge spillovers theories, I extend the study carried out in chapter 2 by running a multivariate analysis based on a conditional logistic regression. Additionally, I perform several sensitivity tests to verify the robustness of the obtained results.

In chapter 4, I acknowledge the suggestion of *DeAngelo* (1981a) that the auditor-client dynamic should be analysed in a multi-period framework, as it is the expected present value of quasi-rents from future contracts which affects the auditor's decisions¹⁹. I argue that using a multi-period approach might help separate the consequences of economic bonding from that of knowledge spillovers: while in the current period, the provision of NAS might increase the ability of the auditor to detect errors, it should not affect his decision to report them, as this depends on the expectation of quasi-rents from future contracts. Therefore, I examine how the current provision of NAS as well as the expectations regarding the future provision of audit and non-audit services affect audit quality, trying to provide further answers to the concerns expressed by the European regulator regarding the provision of certain NAS and the level and structure of the fees paid by audited companies to their statutory auditor.

In chapter 5, I focus on the enforcement system and the ability of the enforcement institutions to detect companies engaging in earnings management. As the enforcement announcements do not clarify whether errors are unintentional or the result of manipulations, they leave an information gap regarding the extent of earnings management among the companies subject to enforcement findings²⁰. Prior studies on the German enforcement system show that enforcement findings cannot be solely attributed to unintentional mistakes and that error companies seem to display incentives and indications of engaging in earnings management²¹. I contribute to this field of research by analysing the relation between enforcement announcements and earnings management using the probability of manipulation (m-score)²². I test whether companies with enforcement findings have, in the error year, a higher probability of manipulation than companies subject to the enforcement

¹⁹ DeAngelo (1981a), p. 115.

²⁰ The wording "enforcement findings" is used to indicate the result of an enforcement examination, which is disclosed in the enforcement announcement. In this study, enforcement findings and enforcement announcements are used interchangeably. I am aware that, in exceptional cases, there might be enforcement findings, which do not get disclosed in an enforcement announcement e.g. when they are not deemed relevant to the financial market, see here e.g. Fiallo/Hecker (2019a), p. 227.

²¹ E.g. Böcking/Gros/Worret (2015), p. 482, Strohmenger (2013), p. 28, Strohmenger (2014), p. 296, and Hoehn/Strohmenger (2013), p. 3.

²² Building on the ratio analysis and the explanatory power of financial statement ratios, Beneish (1999), pp. 24-26, developed a series of indices, based on financial statement data, that added together give the probability of manipulation (m-score) for a certain company in a given year. For more details, see section 5.4.

regime, which have not issued any enforcement announcements, and whether the existence of enforcement announcements can be predicted by using different financial statement ratios. Finally, I focus on the information content of announcements, in terms of error severity and managers' lack of cooperation during the examination, to analyse whether they might indicate a higher probability of manipulation.

2. Ausgewählte Neuregelungen der EU-Verordnung zur Abschlussprüfung und Prüfungsqualität (Teil 1, 2 und 3) – Eine empirische Bestandsaufnahme bei deutschen Unternehmen mit DPR-Fehlerfeststellungen

Abstract: Die Neuregelung der Verordnung (EU) Nr. 537/2014 sollen zur Verbesserung der Qualität der Abschlussprüfung von PIES beitragen. Die in dieser Studie durchgeführte deskriptive Analyse von Unternehmen mit DPR-Fehlerveröffentlichungen und einer Stichprobe von Vergleichsunternehmen ohne Fehlerveröffentlichungen liefert kein Argument für die durch Einführung der black list vorgenommene Verschärfung der Regelungen hinsichtlich des Angebots von Nichtprüfungsleistungen. Teil 2 des Beitrags unterstützt Zweifel daran, dass das cap eine besonders wirksame Maßnahme zur Steigerung der Prüfungsqualität ist. Hinsichtlich der 15%-Honorargrenze ergeben sich Hinweise für die grds. richtige Stoßrichtung der Regelung. Es verbleiben jedoch Zweifel an deren Effektivität. Aus Teil 3 des Beitrags kann kein Argument gegen eine externe Rotation abgeleitet werden.

This chapter has been written together with Renate Hecker and published in Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung (KoR). It is divided into three parts for publication purposes (Parts 1, 2 and 3). Reference to this chapter are made throughout the dissertation based on the publication in the KoR and are cited as Fiallo/Hecker (2019a), Fiallo/Hecker (2019b) and Fiallo/Hecker (2019c).

2.1 Einleitung

Die Finanzkrise 2008 hat auch im Bereich der Abschlussprüfung eine Regulierungswelle ausgelöst, die auf europäischer Ebene in der Richtlinie 2014/56/EU zur Änderung der Richtlinie 2006/43/EG über Abschlussprüfungen von Jahresabschlüssen und konsolidierten Abschlüssen (EU-RL)¹ sowie in der Verordnung (EU) Nr. 537/2014 (EU-VO) über spezifische Anforderungen an die Abschlussprüfung bei Unternehmen von öffentlichem Interesse (PIEs²) vom April 2014 ihren Abschluss fand. Während die zwingenden Vorschriften der EU-VO in den Mitgliedstaaten der EU und damit auch in Deutschland seit 17.06.2016 unmittelbare Gültigkeit entfalten,³ wurden die Mitgliedstaatenwahlrechte der EU-VO und alle Vorschriften der EU-RL, soweit sie unmittelbar prüfungsrelevant sind, vom deutschen Gesetzgeber durch das sog. Abschlussprüfungsreformgesetz (AReG) vom 10.05.2016 umgesetzt und traten fristgerecht zum 17.06.2016 in Kraft. Mit dem Ziel der Verbesserung der Qualität der Abschlussprüfung hat die EU-VO für PIEs u.a. die Erbringung von Nichtprüfungsleistungen (NPL) durch den Abschlussprüfer sowohl hinsichtlich der Kategorien zulässiger NPL als auch im Hinblick auf den erlaubten Umfang (cap, zulässiger Umsatzanteil von Honoraren aus Prüfungs- und Nichtprüfungsleistungen) weiter eingeschränkt. Auch mit der erstmaligen verbindlichen Einführung der sog. externen Rotation soll die Prüfungsqualität verbessert werden.⁴

In der Literatur finden sich nun durchaus Zweifel, ob die konkreten Vorgaben der EU-VO tatsächlich geeignet sind, zur Verbesserung der Abschlussprüfungsqualität beizutragen.⁵ Vor diesem Hintergrund soll in diesem Beitrag für eine Stichprobe von Fällen mit tendenziell niedriger Prüfungsqualität eine deskriptive Auswertung dahingehend erfolgen, inwieweit die durch die EU-VO vorgesehenen Beschränkungen hinsichtlich Art und/oder Umfang von NPL bei diesen Fällen

¹ Die Abkürzung EU-RL wird im Folgenden für die überarbeitete Abschlussprüferrichtlinie verwendet. Die Richtlinie 2014/56/EU zur Änderung der Richtlinie 2006/43/EG wurde im ABIEU Nr. L 158 vom 27.05.2014 S. 196 veröffentlicht.

² PIE steht für Public Interest Entity. Der Begriff umfasst gem. Art. 2 Nr. 13 EU-RL mindestens kapitalmarktorientierte Unternehmen, bei denen die Handelszulassung in einem EU-Staat bereits erfolgt ist, sowie bestimmte Banken und Versicherungen. Im Übrigen besteht für jeden Mitgliedstaat ein Wahlrecht zur Ausweitung dieses Begriffs auf durch den Mindestumfang nicht abgedeckte Unternehmen. Für Deutschland folgt die Abgrenzung der PIE aus §§ 317 Abs. 3a, 340k Abs. 1 Satz 4, 341k Abs. 1 Satz 4 HGB. Aufgrund der Ausübung des Mitgliedstaatenwahlrechts nach Art. 2 Abs. 3 EU-VO sind Genossenschaftsbanken und Sparkassen in Deutschland von bestimmten Vorschriften der EU-VO ausgenommen; vgl. die einschlägigen Vorschriften in Abschn. 4 GenG, sowie § 340k Abs. 3 u. 4 HGB.

³ Vgl. Art. 44 EU-VO. Eine Ausnahme stellt lediglich Art. 16 Abs. 6 EU-VO dar, der erst ab 17.06.2017 in Kraft trat. Die EU-VO wurde im ABIEU Nr. L 158 vom 27.05.2014 S. 77 veröffentlicht.

⁴ Vgl. Erwägungsgründe 7, 8 und 21 der EU-VO.

⁵ Vgl. im Hinblick auf black list und cap z.B. Ratzinger-Sakel/Schönberger, *Accounting in Europe* 2015 S. 62 f.; im Hinblick auf die externe Rotation vgl. z.B. Pott/Schröder/Weckelmann, *KoR* 2014 S. 504.

überhaupt zum Tragen gekommen wären, wenn sie im Untersuchungszeitraum bereits zu beachten gewesen wären. Die Ausübung bzw. Nichtausübung in der EU-VO vorgesehener Mitgliedstaatenwahlrechte im Rahmen des AReG wird dabei jeweils mitberücksichtigt. Betrachtet wird schließlich auch, inwieweit bei diesen Fällen mit tendenziell niedriger Prüfungsqualität eine freiwillige Prüferrotation zu beobachten war. Die Bestandsaufnahme wird jeweils mit dem Befund einer Kontrollstichprobe verglichen.

Da als Surrogatmaß für eine niedrige Prüfungsqualität in dieser Untersuchung zwischen dem 01.07.2005 und dem 31.12.2017 im Bundesanzeiger veröffentlichte Meldungen von Rechnungslegungsfehlern verwendet werden, die von der Deutschen Prüfstelle für Rechnungslegung (DPR) bzw. der BaFin im Rahmen eines sog. Enforcementverfahrens festgestellt wurden, wird in Abschn. 2.2 zunächst der Begriff der Prüfungsqualität erläutert und verdeutlicht, wo die Grenzen des gewählten Surrogatmaßes liegen. Dabei erfolgt für die Stichprobe der Fehlermeldungen auch eine Auswertung der jeweiligen Bestätigungsvermerke für den von der DPR bzw. der BaFin als fehlerhaft eingestuften Abschluss bzw. Lagebericht. Abschn. 2.3 verdeutlicht die genaue Abgrenzung der Stichprobe der Fehlermeldungen sowie der Kontrollstichprobe und erläutert die weiteren für die Untersuchung erhobenen Daten. Die empirische Auswertung hinsichtlich der Kategorien angebotener Prüfungs- und Nichtprüfungsleistungen folgt in Abschn. 2.4. Abschn. 2.5 fasst die wesentlichen Befunde der Bestandsaufnahme in Teil 1 dieses Beitrags zusammen. Der demnächst folgende Teil 2 befasst sich aus empirischer Sicht mit dem Umfang der angebotenen Prüfungs- und Nichtprüfungsleistungen. Der darauf folgende Teil 3 ist der Erhebung der freiwilligen Prüferwechsel gewidmet.

2.2 Fehlerveröffentlichungen nach Abschluss eines Enforcementverfahrens als Surrogatmaß für Prüfungsqualität

Die Bereitstellung glaubwürdiger Informationen durch die Rechnungslegung ist eine entscheidende Voraussetzung für die Funktionsfähigkeit des Kapitalmarkts. Dem Abschlussprüfer kommt dabei eine wichtige Kontrollfunktion zu, weil es seine Aufgabe ist, zu überprüfen, ob der vom Management verantwortete Jahres- bzw. Konzernabschluss und zugehörige Lagebericht den rechtlichen Vorgaben entspricht, also im Wesentlichen frei von Fehlern und damit verlässlich ist.⁶ Ob bzw. inwieweit es dem Prüfer gelingt, Vertrauen in die Rechnungslegung herzustellen, wird

⁶ Vgl. z.B. Wagenhofer/Ewert, Externe Unternehmensrechnung, 3. Aufl. 2015, S. 410.

davon abhängen, wie die Adressaten der Rechnungslegung die Qualität der Abschlussprüfung einschätzen. Unter Prüfungsqualität kann man, einer geläufigen Definition von *DeAngelo* folgend, die vom Markt eingeschätzte Wahrscheinlichkeit dafür verstehen, dass ein Abschlussprüfer einen Fehler in der Rechnungslegung des Mandanten sowohl entdeckt als auch berichtet.⁷ Die Entdeckungswahrscheinlichkeit hängt dabei von der eingesetzten Prüfungstechnologie bzw. der Fachkompetenz des Prüfers ab, wohingegen die bedingte Wahrscheinlichkeit dafür, dass ein Fehler nach seiner Entdeckung auch berichtet wird, vom Grad der Unabhängigkeit des Prüfers vom jeweiligen Mandanten beeinflusst wird.⁸ Grds. ist weder die Vorgehensweise bei der Prüfung noch sind die spezifischen Anreize, denen der Prüfer bei einem bestimmten Mandanten unterliegt, von Außenstehenden, d.h. z.B. von Kapitalmarktteilnehmern, aber auch von Wissenschaftlern, direkt bzw. kostenlos beobachtbar.⁹ Insofern ist man für die Messung von Prüfungsqualität auf die Heranziehung öffentlich verfügbarer Surrogatgrößen angewiesen.¹⁰

Im Folgenden werden als Surrogatmaß für Prüfungsqualität zwischen dem 01.07.2005 und dem 31.12.2017 im Bundesanzeiger veröffentlichte Meldungen von Rechnungslegungsfehlern verwendet, die von der DPR bzw. der BaFin im Rahmen eines sog. Enforcementverfahrens gem. §§ 342b-342e HGB, 106-113 WpHG bei einem kapitalmarktorientierten Unternehmen festgestellt wurden. DPR-Fehlermeldungen als Surrogatmaß für Prüfungsqualität wurden z.B. bereits verwendet bei *Frey/Möller/Weinzierl* (2016) im Hinblick auf den Zusammenhang zwischen dem Wechsel des Abschlussprüfers und der Prüfungsqualität¹¹ sowie bei *Gros* (2016) bzw. im Rahmen einer Robustheitsanalyse bei *Krauß/Pronobis/Zülch* (2015) hinsichtlich des Zusammenhangs zwischen NPL bzw. abnormalen Prüfungsgebühren und Prüfungsqualität.¹²

⁷ Vgl. DeAngelo, *Journal of Accounting and Economics* 1981 S. 186.

⁸ Vgl. DeAngelo, *Journal of Accounting and Economics* 1981 S. 186. Zu beachten ist, dass die Frage der eingesetzten Prüfungstechnologie insofern auch vom Grad der Unabhängigkeit abhängen dürfte, als ein Prüfer, der einen entdeckten Fehler ohnehin nicht berichten würde, naturgemäß auch keinen Anreiz hat, zunächst besonders sorgfältig zu prüfen; vgl. DeAngelo, *Journal of Accounting and Economics* 1981 S. 116, Fn. 3.

⁹ Vgl. DeAngelo, *Journal of Accounting and Economics* 1981 S. 186.

¹⁰ Vgl. zu einem Überblick über bisher in der Prüfungsforschung vornehmlich verwendete Surrogate für Prüfungsqualität Quick/Schmidt/Simons, *WPg* 2016 S. 196 ff.

¹¹ Vgl. Frey/Möller/Weinzierl, *KoR* 2016 S. 563 (571), wo analog zu Teil 3 des vorliegenden Beitrags insb. auch Prüferwechsel vor dem Jahr des fehlerhaften Abschlusses betrachtet werden.

¹² Vgl. Krauß/Pronobis/Zülch, *Journal of Business Economics* 2015 S. 71-75. Hinsichtlich Gros (2016) ist allerdings anzumerken, dass dort aufgrund des Zugangs zu nicht öffentlich verfügbaren Informationen der BaFin als Surrogatmaß für hohe Prüfungsqualität Enforcementverfahren verwendet werden, bei denen keine Fehler festgestellt wurden; vgl. Gros, *International Journal of Economics and Accounting* 2016 S. 219.

Da bei der genauen Abgrenzung der Stichprobe in Abschn. 2.3 u.a. alle Fehlermeldungen ausgeschlossen werden, die sich lediglich auf einen verkürzten und damit nicht prüfungspflichtigen Abschluss bzw. Lagebericht beziehen, werden durchweg nur Fehler erfasst, die nach abgeschlossener Abschlussprüfung und Veröffentlichung des Abschlusses bzw. Lageberichts noch in der Rechnungslegung enthalten sind und von der DPR bzw. der BaFin als wesentlich angesehen werden.¹³ Die erhobenen Fehlermeldungen sind jedoch aus dreierlei Gründen kein perfektes Maß für eine schlechte Qualität der Abschlussprüfung.

So kann zunächst einmal dann nicht von einem Fehler des Abschlussprüfers gesprochen werden, wenn dieser die Rechnungslegungsfehler bei seiner Prüfung erkannt und deshalb den Bestätigungsvermerk bereits eingeschränkt oder versagt hat. Deshalb wird die in der Folge für eine niedrige Prüfungsqualität verwendete Surrogatgröße dahingehend präzisiert, dass alle Fehlermeldungen aus der endgültigen Stichprobe ausgeschlossen werden, bei denen der Abschlussprüfer zuvor keinen uneingeschränkten Bestätigungsvermerk erteilt hat. Nur bei einem uneingeschränkten Bestätigungsvermerk kann davon ausgegangen werden, dass der Abschlussprüfer den Fehler zuvor nicht entdeckt oder zumindest nicht als wesentlichen Fehler eingeschätzt bzw. ggf. aus mangelnder Unabhängigkeit vom Mandanten jedenfalls nicht berichtet hat. Wie Tab. 2.1 zu entnehmen ist, verkleinert sich die zunächst verbleibende Stichprobe von 173 Fehlermeldungen dabei jedoch nur um 22 Fälle bzw. 12.7%, weil lediglich drei Versagungsvermerke und 19 eingeschränkte Bestätigungsvermerke vorliegen.¹⁴

¹³ Vgl. zur Beachtung der Wesentlichkeit bei Fehlerfeststellung bzw. -veröffentlichung Scheffler, IRZ 2006 S. 17 f.; DPR, 10 Jahre Bilanzkontrolle in Deutschland (2005 bis 2015), 2015, S. 49; BaFin, Emittentenleitfaden der BaFin, 5. Aufl. 2018, Modul A, S. 8 f.

¹⁴ 25 der insgesamt 151 uneingeschränkten Bestätigungsvermerke sind durch einen Hinweis ergänzt. Dies gilt auch für sechs der 19 eingeschränkten Bestätigungsvermerke. In Tab. 2.1 wurde auf einen Ausweis dieser sechs Fälle in einer separaten Spalte verzichtet.

Tab. 2.1: Ergebnisse der Abschlussprüfung der fehlerhaften Abschlüsse und deren Verteilung über die Zeit

Jahr des Bilanzstichtags des fehlerhaften Abschlusses	Versagungsvermerk	Eingeschränkter Bestätigungsvermerk	Uneingeschränkter Bestätigungsvermerk mit Hinweis	Uneingeschränkter Bestätigungsvermerk	Summe	Anteil
2004	0 (0,0%)	1 (25,0%)	0 (0,0%)	3 (75,0%)	4	2,3%
2005	0 (0,0%)	7 (23,3%)	4 (13,3%)	19 (63,3%)	30	17,3%
2006	0 (0,0%)	1 (3,8%)	6 (23,1%)	19 (73,1%)	26	15,0%
2007	0 (0,0%)	2 (14,3%)	2 (14,3%)	10 (71,4%)	14	8,1%
2008	0 (0,0%)	2 (9,5%)	5 (23,8%)	14 (66,7%)	21	12,1%
2009	1 (4,8%)	0 (0,0%)	3 (14,3%)	17 (81,0%)	21	12,1%
2010	1 (9,1%)	2 (18,2%)	0 (0,0%)	8 (72,7%)	11	6,4%
2011	0 (0,0%)	1 (12,5%)	0 (0,0%)	7 (87,5%)	8	4,6%
2012	0 (0,0%)	1 (6,3%)	2 (12,5%)	13 (81,3%)	16	9,2%
2013	1 (16,7%)	0 (0,0%)	0 (0,0%)	5 (83,3%)	6	3,5%
2014	0 (0,0%)	2 (22,2%)	2 (22,2%)	5 (55,6%)	9	5,2%
2015	0 (0,0%)	0 (0,0%)	1 (14,3%)	6 (85,7%)	7	4,0%
Summe	3	19	25	126	173	
Anteil	1,7%	11,0%	14,5%	72,8%		100%

Die Zahlen in Klammern geben jew. den prozentualen Anteil an der Gesamtzahl der Fälle des in der jeweiligen Zeile betrachteten Jahres an.

Beschränkt man sich somit auf Fehlerbekanntmachungen, bei denen der Abschlussprüfer zuvor für den als fehlerhaft eingestuften Konzernabschluss bzw. -lagebericht einen uneingeschränkten Bestätigungsvermerk erteilt hat, so hat man dennoch nach wie vor noch kein perfektes Maß für Abschlussprüfungsqualität. Die DPR bzw. die BaFin prüft nämlich nicht in jedem Jahr alle dem Enforcement unterliegenden Unternehmen¹⁵ und legt ihren Fokus aufgrund der von der DPR jährlich im Voraus festgelegten Prüfungsschwerpunkte¹⁶ oder aufgrund des Anlasses der Prüfung regelmäßig auch jeweils nur auf einen Teilbereich des Abschlusses bzw. Lageberichts.¹⁷ Darüber hinaus hat sie grds. auch keinen so tiefen Einblick in das zu prüfende Unternehmen wie der

¹⁵ Vgl. DPR, Tätigkeitsbericht 2005, S. 13. Den Grundsätzen für die Stichprobenprüfung der DPR ist zu entnehmen, dass gewährleistet werden soll, dass alle Unternehmen, die dem Enforcement unterliegen, mindestens alle acht bis zehn Jahre einmal von der DPR überprüft werden, alle Unternehmen des DAX, MDAX, SDAX und TecDAX sogar alle vier bis fünf Jahre. Insofern sollten bis 31.12.2017 eigentlich alle Unternehmen, die dem Enforcement unterliegen, mindestens bereits einmal der Überprüfung unterlegen haben. Dies gilt auch für noch nicht lange gelistete Unternehmen, weil ein erstmaliges Listing regelmäßig dazu führen dürfte, dass ein Unternehmen im Rahmen der Stichprobenprüfung bereits auf der ersten Stufe bei der risikoorientierten Auswahl ausgewählt wird; vgl. DPR, Grundsätze für die stichprobenartige Prüfung gem. § 342b Abs. 2 Satz 3 Nr. 3 HGB vom 05.09.2005, 20.04.2009 und 15.12.2016. Ausnahmen sind aber natürlich denkbar.

¹⁶ Vgl. DPR, a.a.O. (Fn. 13), S. 41.

¹⁷ Vgl. DPR, Tätigkeitsbericht 2005, S. 12 f.

Abschlussprüfer¹⁸ und die BaFin kann gem. § 109 Abs. 2 Satz 2 oder 3 WpHG in seltenen Fällen auf eine Veröffentlichungsanordnung verzichten.¹⁹ Deshalb kann nicht davon ausgegangen werden, dass im Rahmen des Enforcements alle wesentlichen Fehler festgestellt und veröffentlicht werden, die vom zuständigen Abschlussprüfer ggf. zuvor fälschlicherweise nicht berichtet wurden. Diese Überlegung spricht allerdings nicht gegen die Interpretation der Fehlermeldungen als Hinweis auf eine in diesen Fällen tendenziell niedrige Prüfungsqualität, sondern impliziert, dass eine nicht vorhandene Fehlermeldung nicht zwingend auf eine hohe Qualität der Prüfung hindeuten muss,²⁰ was somit insb. eine Einschränkung bei der Festlegung der Kontrollstichprobe darstellt.

Was die geforderte Treffsicherheit im Hinblick auf die Aufdeckung wesentlicher Fehler durch den Abschlussprüfer betrifft, so sehen die einschlägigen Vorgaben der International Standards on Auditing (ISA)²¹ genau wie die bisher in Deutschland zu beachtenden Prüfungsstandards des IDW (IDW PS) im Rahmen der Anwendung eines risikoorientierten Prüfungsansatzes lediglich eine hinreichende Sicherheit für die getroffenen Prüfungsaussagen vor. Eine gewisse, hinreichend kleine Irrtumswahrscheinlichkeit wird also akzeptiert.²² In diesem Sinne muss dann nicht jeder im Nachhinein entdeckte und berichtete wesentliche Fehler, also nicht jede im Folgenden betrachtete Fehlermeldung, bereits zwingend auf ein Prüferversagen im rechtlichen Sinne hindeuten.²³ Je effektiver und damit anspruchsvoller die vom Regulator gesetzte Benchmark hinsichtlich anzuwendender Prüfungstechnologie und einzuhaltender Unabhängigkeitsregeln wäre bzw. je kleiner die tolerierte Irrtumswahrscheinlichkeit angesetzt würde, desto weniger käme naturgemäß diese Einschränkung des hier verwendeten Surrogatmaßes für Prüfungsqualität zum Tragen.²⁴ Im Übrigen ist im Hinblick auf alternativ denkbare Maße zu berücksichtigen, dass das bisher in der

¹⁸ Vgl. Beyhs/Kühne/Zülch, WPg 2012 S. 654 f.

¹⁹ Zur Auslegung dieser Regelung vgl. BaFin, a.a.O. (Fn.13), S. 8 f. Bei allen von Juli 2005 bis 31.12.2017 von der BaFin abgeschlossenen Enforcementverfahren hat die BaFin nur in acht Fällen auf eine Fehlerveröffentlichung verzichtet; vgl. BaFin, Jahresbericht 2013, S. 181; BaFin, Jahresbericht 2014, S. 236; BaFin, Jahresbericht 2015, S. 251; BaFin, Jahresbericht 2016, S. 192; sowie BaFin, Jahresbericht 2017, S. 149.

²⁰ Auf diese Einschränkung wird z.B. bereits hingewiesen bei Quick/Schmidt/Simons, WPg 2016 S. 198, sowie im Zusammenhang mit SEC Enforcement Actions bei Francis, *Auditing: A Journal of Practice and Theory* 2011 S. 128.

²¹ Die Beachtung der ISA sieht Art. 26 EU-RL vorbehaltlich ihres Endorsement zukünftig für Abschlussprüfungen verpflichtend vor; vgl. für Deutschland insoweit § 317 Abs. 5 und 6 HGB; vgl. für PIE zusätzlich Art. 9 EU-VO.

²² Vgl. z.B. Schmidt/Almeling, in: Grottel u.a. (Hrsg.), Beck Bil-Komm., 11. Aufl. 2018, § 317 HGB Rdn. 10, mit Verweis auf IDW PS 200, Tz. 9 und 24 f., sowie im Hinblick auf die internationalen Prüfungsstandards ISA 200, Introduction, Rdn. 5.

²³ Auf diese Einschränkung wird z.B. hingewiesen bei DeFond/Zhang, *Journal of Accounting and Economics* 2014 S. 284. Vgl. zum „legal view of audit quality“ Francis, *Auditing: A Journal of Practice and Theory* 2011 S. 127.

²⁴ Vgl. hierzu auch DeFond/Zhang, *Journal of Accounting and Economics* 2014 S. 279, wo höhere Prüfungsqualität als größere Sicherheit für hohe Rechnungslegungsqualität definiert wird.

Prüfungsforschung häufig verwendete Surrogatmaß für Prüfungsqualität in Form sog. diskretionärer Periodenabgrenzungen spezifische Einschränkungen aufweist, die für das hier gewählte Maß weniger einschlägig erscheinen.²⁵

2.3 Datenbasis und Abgrenzung der Stichprobe

Wie in Abschn. 2.2 erläutert, werden im Folgenden als Surrogatmaß für eine niedrige Prüfungsqualität zunächst alle seit Beginn der Enforcementverfahren am 01.07.2005 bis zum 31.12.2017 im Bundesanzeiger veröffentlichten Fehlermeldungen erhoben. Nach diversen Bereinigungen, die Tab. 2.2 zu entnehmen sind, verbleibt die endgültige Stichprobe von 151 Fehlermeldungen,²⁶ die sich ausschließlich auf IFRS-Konzernabschlüsse und/oder die zugehörigen Lageberichte beziehen.

Tab. 2.2: Abgrenzung der Stichprobe

Von 01.07.2005 bis 31.12.2017 im Bundesanzeiger veröffentlichte Fehlerfeststellungen	258^{a)}
abzüglich korrigierter Fehlermeldungen zu demselben Sachverhalt	17
abzüglich Fehlermeldungen nur hinsichtlich verkürzter Abschlüsse	30
abzüglich Fehlermeldungen hinsichtlich ausländischer Unternehmen	19
abzüglich Fehlermeldungen nur hinsichtlich Jahresabschlüssen (nach HGB)	11
abzüglich Fehlermeldungen hinsichtlich Konzernabschlüssen nach US-GAAP oder HGB)	8
Stichprobe vor Auswertung der Bestätigungsvermerke	173
abzüglich Fehlermeldungen, bei denen ein Versagungsvermerk erteilt wurde	3
abzüglich Fehlermeldungen, bei denen der Bestätigungsvermerk eingeschränkt wurde	19
Stichprobe untersuchter Fehlerveröffentlichungen	151
In einigen wenigen Fällen wurden Fehlerfeststellungen für ein Unternehmen zu Abschlüssen mit unterschiedlichen Abschlussstichtagen, für die normalerweise getrennte Fehlerbekanntmachungen veröffentlicht werden, in einer Feststellung zusammengefasst. Um insoweit eine einheitliche Vorgehensweise zu gewährleisten, wird pro Abschlussstichtag eine separate Fehlermeldung gezählt.	

Der Auswertung der Stichprobe von Fällen mit tendenziell niedriger Prüfungsqualität soll im Folgenden eine analoge Auswertung einer Kontrollstichprobe von Fällen mit tendenziell hoher Prüfungsqualität gegenübergestellt werden. Bei der Auswahl der Unternehmen für die Kontrollstichprobe wird zu jeder Fehlermeldung eines Unternehmens der Stichprobe genau ein bezogen auf das jeweilige Jahr der fehlerhaften Rechnungslegung möglichst gut vergleichbares

²⁵ Vgl. zu dieser Einschätzung ebenso Quick/Schmidt/Simons, WPg 2016 S. 196-198.

²⁶ Fehlerbekanntmachungen, die sich auf dasselbe Unternehmen, aber auf Abschlüsse unterschiedlicher Jahre beziehen, verbleiben in der Stichprobe. Fehlermeldungen für mehr als zwei unterschiedliche Geschäftsjahre sind jedoch für kein Unternehmen in der in Tab. 2.2 abgegrenzten Stichprobe untersuchter Fehlerveröffentlichungen enthalten.

Unternehmen (1-1-Match) gesucht,²⁷ das im gesamten Untersuchungszeitraum von 01.07.2005 bis 31.12.2017 keine aus einem Enforcementverfahren resultierende Fehlerbekanntmachung im Bundesanzeiger vorgenommen hat und folgende weitere Kriterien erfüllt:²⁸ Es ist ein deutsches Unternehmen, das derselben Branche angehört wie das Stichprobenunternehmen.²⁹ Es erstellt im Jahr der fehlerhaften Rechnungslegung des Stichprobenunternehmens ebenfalls einen IFRS-Konzernabschluss³⁰ und gehört in diesem Jahr zu den Unternehmen, die im Rahmen eines Enforcementverfahrens von der DPR bzw. der BaFin geprüft werden könnten.³¹ Es weist schließlich eine ähnliche Unternehmensgröße wie das Stichprobenunternehmen auf.³²

Im Hinblick auf die Kontrollstichprobe ist zu beachten, dass mittels der Auswahlkriterien für die Vergleichsunternehmen eine hohe Prüfungsqualität bei diesen Unternehmen im jeweiligen Jahr nicht garantiert werden kann, weil, wie im vorigen Abschn. 2.2 ausführlich erläutert, eine nicht

²⁷ Ein 1-1-Match findet sich in mehreren bisherigen Studien zum deutschen Enforcement; vgl. z.B. Strohmenger, *Journal of International Financial Management & Accounting* 2014 S. 281, sowie Ebner/Hottmann/Zülch, *Corporate Ownership & Control* 2017 S. 128.

²⁸ Die Auswahl der Vergleichsunternehmen erfolgt dabei aus allen im Dezember 2015 in Thomson Reuters Eikon enthaltenen Unternehmen mit Sitz in Deutschland, die an einer inländischen Börse notierte Wertpapiere emittiert haben. Für die Überprüfung der Erfüllung der Matchingkriterien wird überwiegend diese Datenbank herangezogen.

²⁹ Die Branchenzuordnung wird nach den Sektoren des Global Industry Classification Standard (GICS) vorgenommen.

³⁰ Eine Ausnahme besteht insofern lediglich für zwei Vergleichsunternehmen. Bei einem Unternehmen handelt es sich um einen Fall aus dem Jahr 2004, der ausschließlich in die Auswertung in Teil 3, Abschn. 2.10, eingeht. Bei dem zweiten Fall aus dem Jahr 2005 waren weder ein Einzel- noch ein Konzernabschluss zu finden. Es konnten jedoch die Vergleichszahlen nach IFRS sowie die Prüferhonorare für 2005 aus dem IFRS-Konzernabschluss des Jahres 2006 entnommen werden. Dieser Fall geht ausschließlich in die Auswertungen in Abschn. 2.4 und in Teil 3, Abschn. 2.10, ein.

³¹ Ob die DPR bzw. die BaFin im jeweiligen Jahr das Recht gehabt hätte, die Rechnungslegung des jeweiligen Unternehmens im Rahmen eines Enforcementverfahrens zu prüfen, wird näherungsweise anhand der einschlägigen Listen bestimmt, die die BaFin hierfür jährlich veröffentlicht; vgl. hierzu <http://hbfm.link/4832>, letzter Abruf: 30.05.2018. Die älteren Listen wurden auf Anfrage unmittelbar von der BaFin bezogen.

³² Gewählt wird dabei jeweils das Unternehmen, dessen Konzernbilanzsumme im jeweiligen Jahr des Fehlers die jeweils geringste absolute Differenz zur Konzernbilanzsumme des Stichprobenunternehmens aufweist, vorausgesetzt, dass dieses Unternehmen für das jeweilige Jahr nicht bereits in der Kontrollstichprobe enthalten ist. Sollte Letzteres der Fall sein, wird stattdessen das Unternehmen gewählt, das zur zweitkleinsten absoluten Konzernbilanzsummendifferenz führt usw. Hinsichtlich einer Unternehmens-Jahr-Beobachtung wird somit bei der Erstellung der Kontrollstichprobe ohne Zurücklegen gearbeitet, sodass keine Unternehmens-Fehlerjahr-Kombination mehrfach in der Kontrollstichprobe enthalten sein kann. Für unterschiedliche Jahre kann hingegen, analog zur Vorgehensweise bei der Abgrenzung der Stichprobe der Fehlermeldungen, dasselbe Unternehmen erneut in die Kontrollstichprobe aufgenommen werden. Dies führt für die vorliegende Stichprobe dazu, dass ein Unternehmen maximal dreimal in der Kontrollstichprobe vertreten ist. Letztere Vorgehensweise verbessert die Qualität des Match im Hinblick auf die Vergleichbarkeit der Unternehmensgröße, die weiter sinken würde, wenn nicht nur völlig identische Beobachtungen (d.h. gleiche Unternehmens-Jahr-Kombinationen), sondern auch Beobachtungen gleicher Unternehmen in unterschiedlichen Jahren, die möglicherweise ebenfalls nicht völlig unkorreliert sind, ausgeschlossen würden. Letztlich besteht hier also ein trade-off zwischen der Qualität des Match und der Unabhängigkeit zwischen den Beobachtungen; vgl. hierzu allgemein z.B. Shipman/Swanquist/Whited, *The Accounting Review* 2017 S. 217, sowie Stuart, *Statistical Science* 2010 S. 9.

vorhandene Fehlermeldung nicht zwingend auf eine hohe Qualität der Prüfung hindeuten muss. Insofern sind die fehlenden Fehlermeldungen der Unternehmen der Kontrollstichprobe lediglich als ein ebenfalls gewissen Einschränkungen unterliegendes Surrogatmaß für eine hohe Prüfungsqualität anzusehen.

Um Aussagen hinsichtlich Art und Umfang von Prüfungs- und Nichtprüfungsleistungen für die Unternehmen der Stichprobe und die Vergleichsunternehmen der Kontrollstichprobe treffen zu können, wurden zunächst für das Jahr, auf das sich der gemeldete Fehler bezieht, jeweils die gem. § 314 Abs. 1 Nr. 9 HGB für den Konzernabschlussprüfer anzugebenden Honorare händisch aus den Anhängen der Konzernabschlüsse von Stichproben- und Vergleichsunternehmen erhoben. Da diese Pflicht zur Angabe der Honorare gem. Art. 58 Abs. 3 Satz 1 EGHGB erstmals für Geschäftsjahre greift, die nach dem 31.12.2004 beginnen, müssen mangels Daten alle Fälle, die sich auf das Kalenderjahr 2004 oder auf ein in 2004 beginnendes vom Kalenderjahr abweichendes Geschäftsjahr 2004/05 beziehen, von den Auswertungen im folgenden Abschn. 2.4 und in Teil 2 des Beitrags ausgeschlossen werden.³³

Für Teil 2, Abschn. 2.7.2, wird für alle Fälle der Stichprobe und Kontrollstichprobe noch das insgesamt von allen Mandanten des jeweiligen Konzernabschlussprüfers erzielte Honorar erhoben, um den Anteil der jeweiligen PIE am Gesamtumsatz des Prüfers abzuschätzen. Die Umsatzdaten werden dem jeweiligen Transparenzbericht des Prüfers entnommen, in dem gem. § 55c Abs. 1 Satz 3 Nr. 3 WPO a.F. von Wirtschaftsprüfungsgesellschaften, die im Jahr mindestens eine Abschlussprüfung einer PIE gem. § 319a Abs. 1 Satz 1 HGB durchführen, der gem. § 285 Nr. 17 HGB nach Honoraren aufgeschlüsselte Gesamtumsatz anzugeben war. Der Pflicht, jährlich spätestens zum 31.03. einen Transparenzbericht zu veröffentlichen, mussten Abschlussprüfer einer PIE erstmals zum 31.03.2008³⁴ nachkommen, sodass aufgrund noch nicht verfügbarer Transparenzberichte in Teil 2, Abschn. 2.7.2, weitere Fälle von der Auswertung ausgeschlossen werden müssen, die sich auf Geschäftsjahre zu Beginn des Auswertungszeitraums beziehen.³⁵

³³ Die Stichprobe und damit auch die Kontrollstichprobe verringern sich deshalb um jeweils sieben Fälle.

³⁴ Vgl. WPK, WPK-Magazin 1/2009 S. 4.

³⁵ Seit Inkrafttreten der EU-VO zum 17.06.2016 folgt die Verpflichtung zur Veröffentlichung eines Transparenzberichts für Abschlussprüfer bzw. Prüfungsgesellschaften von PIE unmittelbar aus Art. 13 Abs. 1 EU-VO. Gem. Verlautbarung Nr. 2 der Abschlussprüferaufsichtsstelle (APAS) vom 07.03.2017 zur Veröffentlichung eines Transparenzberichts wird es die APAS trotz insoweit nicht vorhandener Übergangsregelung zu Art. 13 EU-VO nicht beanstanden, wenn für Geschäftsjahre, die vor dem 17.06.2016 begonnen haben, noch einmal ein Transparenzbericht

Was nun die Interpretation der durch § 314 Abs. 1 Nr. 9 HGB vorgeschriebenen Angaben zum Honorar des Konzernabschlussprüfers betrifft, so ist zu berücksichtigen, dass die Regelung durch das Bilanzrechtsmodernisierungsgesetz (BilMoG) in mehrfacher Hinsicht mit Wirkung für nach dem 31.12.2008 beginnende Geschäftsjahre geändert wurde.³⁶ Dies ist für die vorliegende Bestandsaufnahme insofern relevant, als sowohl Fälle vor als auch nach Inkrafttreten des BilMoG betrachtet werden. Die Änderungen der Regelung betreffen im Wesentlichen den Begriff des Prüfers, die Abgrenzung der Leistungsempfänger im Konzern, die zeitliche Abgrenzung des Gesamthonorars und die Aufschlüsselung des insgesamt anzugebenden Honorars.

Was die Abgrenzung des Prüfers betrifft, so wird in § 314 Abs. 1 Nr. 9 HGB sowohl vor als auch nach dem BilMoG zwar auf den Abschlussprüfer des Konzernabschlusses Bezug genommen. Bis zum BilMoG wurde diese Regelung aber überwiegend so ausgelegt, dass nicht nur das Gesamthonorar des gesetzlichen Konzernabschlussprüfers, sondern auch die Honorare von dessen gem. § 271 Abs. 2 HGB verbundenen Unternehmen mitanzugeben wären.³⁷ Die Honorare von Mitgliedern des (internationalen) Verbunds bzw. Netzwerks des Konzernabschlussprüfers waren indessen nicht berichtspflichtig und sollten bei freiwilliger Angabe durch einen Davon-Vermerk kenntlich gemacht werden.³⁸ Während Letzteres unverändert bis heute gilt,³⁹ wird mit Bezug auf die Gesetzesbegründung seit dem BilMoG hingegen ganz überwiegend davon ausgegangen,⁴⁰ dass der Begriff des Prüfers enger zu fassen ist und nur die Honorare des Konzernabschlussprüfers anzugeben sind, wohingegen die Honorare der mit ihm verbundenen Unternehmen lediglich freiwillig angegeben werden können.⁴¹ Immerhin empfiehlt der IDW RS HFA 36 aber, die Angabe der Honorare für verbundene Unternehmen des Prüfers auch nach Inkrafttreten des BilMoG weiterhin vorzunehmen.⁴² Zusätzlich können seit dem BilMoG separat oder mit Davon-Vermerk

nach § 55c WPO a.F. erstellt wird. Damit konnten mit Duldung der zuständigen Prüferaufsicht zum 31.03.2017 noch einmal Transparenzberichte nach altem Recht erstellt werden.

³⁶ Vgl. Art. 66 Abs. 2 Satz 1 EGHGB.

³⁷ Vgl. z.B. Sultana/Willeke, StuB 2005 S. 953 f.; Bischof, WPg 2006 S. 707 f.; Lenz/Möller/Höhn, BB 2006 S. 1788; IDW RH HFA 1.006 (2005), Rdn. 4-6 i.V.m. Rdn.16; a.A. Ellrott, in: Ellrott u.a. (Hrsg.), Beck Bil-Komm., 6. Aufl. 2006, § 314 HGB Rdn. 91 i.V.m. § 285 HGB Rdn. 269.

³⁸ Vgl. IDW RH HFA 1.006 (2005), Rdn. 5.

³⁹ Vgl. IDW RS HFA 36 n.F., Rdn. 6.

⁴⁰ Vgl. Gesetzentwurf der Bundesregierung zum BilMoG, BT-Drucks. Nr. 16/10067 vom 30.07.2008, S. 70.

⁴¹ Vgl. IDW RS HFA 36, Rdn. 5-7 i.V.m. Rdn. 19, Köhler/Marten/Ratzinger/Wagner, ZfB 2010 S. 8 f., Grottel, in: Grottel u.a. (Hrsg.), Beck Bil-Komm., 11. Aufl. 2018, § 314 HGB Rdn. 160 i.V.m. § 285 HGB Rdn. 513.

⁴² Ein Davon-Vermerk wird insoweit aber nicht explizit verlangt; vgl. IDW RS HFA 36, Rdn. 7 i.V.m. Rdn. 19. Anders insoweit Leidner/Lenz, DBW 2013 S. 383, anders auch der erst für Geschäftsjahre, die nach dem 31.12.2016 beginnen, geltende IDW RS HFA 36 n.F., wo in Rdn. 7 für diesen Fall immerhin eine Davon-Angabe empfohlen wird.

auch Leistungen anderer Abschlussprüfer von in den Konzernabschluss einbezogenen Unternehmen angegeben werden, um die durch das BilMoG eingeführte Befreiungsvorschrift im Einzelabschluss gem. § 285 Nr. 17 HGB in Anspruch zu nehmen.⁴³

Was die Abgrenzung der Leistungsempfänger im Konzern betrifft, so waren vor Inkrafttreten des BilMoG nur Honorare für Leistungen an das Mutterunternehmen sowie alle im jeweiligen Konzernabschluss vollkonsolidierten Tochterunternehmen anzugeben.⁴⁴ Nach Inkrafttreten des BilMoG sollen zusätzlich auch Honorare für Leistungen an quotale konsolidierte Gemeinschaftsunternehmen einzubeziehen sein, wobei eine lediglich quotale Einbeziehung empfohlen wird.⁴⁵

Im Hinblick auf die zeitliche Abgrenzung des Gesamthonorars hat sich der Wortlaut des § 314 Abs. 1 Nr. 9 HGB durch das BilMoG geändert. Während zuvor „das im Geschäftsjahr als Aufwand erfasste Honorar“ anzugeben war, ist seit dem BilMoG auf das „für das Geschäftsjahr berechnete Gesamthonorar“ abzustellen. Dies impliziert insb, dass seit dem BilMoG auch Honorare, die als Anschaffungsnebenkosten gem. § 255 Abs. 1 Satz 2 HGB aktiviert werden, mitanzugeben sind.⁴⁶

Was schließlich die Aufschlüsselung des Gesamthonorars in die vier in § 314 Abs. 1 Nr. 9 HGB vorgesehenen Kategorien betrifft, so ist man bei Zweifeln hinsichtlich der inhaltlichen Zuordnung auch hier auf Gesetzesmaterialien, Kommentierungen und IDW-Verlautbarungen angewiesen und dürfte ein gewisser Gestaltungsspielraum bestehen, der insb. in den Anfangsjahren der Honorarveröffentlichungspflicht besonders groß gewesen sein dürfte.⁴⁷ Seit dem BilMoG sind die anzugebenden Honorare in die vier Kategorien Abschlussprüfungsleistungen (APL), andere Bestätigungsleistungen (aBL), Steuerberatungsleistungen (StBL) und sonstige Leistungen (sL) aufzuschlüsseln. Eine Änderung von Bezeichnung und Inhalt der Kategorie hat der Gesetzgeber

⁴³ Vgl. IDW RS HFA 36, Rdn. 20, sowie Grottel, a.a.O. (Fn. 41), § 314 HGB Rdn. 162 i.V.m. § 285 HGB Rdn. 520 f. Sofern es sich bei diesen Abschlussprüfern nicht um verbundene Unternehmen bzw. Mitglieder des Netzwerks des Konzernabschlussprüfers handelt, werden die für die folgende Auswertung erhobenen Honorare vorab bereits jeweils um diese Honorarbestandteile bereinigt.

⁴⁴ Vgl. IDW RH HFA 1.006, Rdn. 12-15, sowie Ellrott, a.a.O. (Fn. 37), § 314 HGB Rdn. 91.

⁴⁵ Vgl. IDW RS HFA 36, Rdn. 19 und 21, sowie Grottel in: Grottel u.a. (Hrsg.), Beck Bil-Komm., 10. Aufl. 2016, § 314 HGB Rdn. 161. IDW RS HFA 36 n.F., Rdn. 19, sieht die lediglich quotale Einbeziehung in diesen Fällen inzwischen verpflichtend vor. Die Neufassung dieses IDW RS gilt jedoch erst für in der vorliegenden Studie nicht mehr relevante Geschäftsjahre, die nach dem 31.12.2016 beginnen; vgl. IDW RS HFA 36 n.F., Rdn. 4a.

⁴⁶ Vgl. IDW RS HFA 36, Tz. 21 i.V.m. Tz. 8, Grottel, a.a.O. (Fn. 41), § 285 HGB Rdn. 503-508 i.V.m. § 314 HGB Rdn. 161, sowie Sattler, Vereinbarkeit von Abschlussprüfung und Beratung, 2011, S. 115 f.

⁴⁷ Vgl. Zimmermann, KoR 2006 S. 274 f.

dabei bei der zweiten Kategorie vorgenommen, die vor dem BilMoG mit „sonstige Bestätigungs- oder Bewertungsleistungen“ überschrieben war. Bewertungsleistungen, die durch den Konzernabschlussprüfer erbracht werden,⁴⁸ sind seit dem BilMoG in der vierten Kategorie auszuweisen, die nach wie vor mit „sonstige Leistungen“ überschrieben ist und die Funktion eines Auffangtatbestands erfüllt, der alle Honorare erfasst, die nicht den drei anderen Kategorien zuzuordnen sind.⁴⁹

2.4 Untersuchungsergebnisse – Ist die Verschärfung des Verbots von Nichtprüfungsleistungen unter dem Gesichtspunkt der Prüfungsqualität sinnvoll?

Art. 5 Abs. 1 EU-VO führt eine Liste von verbotenen NPL (black list) auf, die grds. zwischen dem Beginn des Prüfungszeitraums und der Abgabe des Bestätigungsvermerks vom Prüfer des (Konzern-)Abschlusses nicht erbracht werden dürfen. Verboten sind danach jeweils alle oder auch nur bestimmte Leistungen in den folgenden Bereichen: Steuerberatung, Unternehmensführung bzw. -entscheidungen, Buchhaltung, Rechnungslegung sowie Abschlusserstellung, Lohn- und Gehaltsabrechnung, Gestaltung und Umsetzung interner Kontroll- oder Risikomanagementsysteme im Zusammenhang mit Finanzinformationen⁵⁰, Bewertung, Recht, interne Revision, Finanzierung bzw. Kapitalstruktur und -ausstattung sowie Anlagestrategie, Werbung für, Handel mit oder Zeichnung von Aktien des jeweiligen Mandanten sowie Personal.

Das Verbot gilt auf der Prüferseite für den Abschlussprüfer der PIE und jedes Mitglied seines Netzwerks und betrifft auf der Seite der Leistungsempfänger alle Leistungen, die direkt oder indirekt für das geprüfte Unternehmen und ggf. dessen Mutterunternehmen sowie dessen Tochterunternehmen mit Sitz in der EU erbracht werden. Erbringen Mitglieder des

⁴⁸ Eigenständige Bewertungsleistungen sind aber bereits ohne Berücksichtigung der einschlägigen Neuregelungen durch die EU-VO gem. § 319 Abs. 3 Satz 1 Nr. 3d) i.V.m. Abs. 5 HGB seit dem BilReG in Deutschland für alle (Konzern-)Abschlussprüfer ohnehin nur noch eingeschränkt zulässig, nämlich dann, wenn sie sich nicht wesentlich auf den zu prüfenden Abschluss auswirken oder die Tätigkeit lediglich von untergeordneter Bedeutung ist.

⁴⁹ Vgl. z.B. Grottel, a.a.O. (Fn. 41), § 285 HGB Rdn. 519 i.V.m. § 314 HGB Rdn. 164, sowie IDW (Hrsg.), WPH Edition, WP Handbuch, Wirtschaftsprüfung & Rechnungslegung, 16. Aufl. 2019, Kap. F, Rdn. 1137 i.V.m. IDW RS HFA 36, Rdn. 15 i.V.m. Rdn. 19.

⁵⁰ Die Gestaltung und Umsetzung interner Kontroll- oder Risikomanagementverfahren, die bei der Erstellung und/oder Kontrolle von Finanzinformationen oder Finanzinformationstechnologiesystemen zum Einsatz kommen, ist dabei nach Art. 5 Abs. 1 Buchst. b) EU-VO auch für das dem Beginn des Prüfungszeitraums unmittelbar vorausgehende Geschäftsjahr nicht gestattet.

Prüfernetzwerks Leistungen für Tochterunternehmen mit Sitz außerhalb der EU, kommen hingegen die etwas weniger restriktiven Vorschriften des Art. 5 Abs. 5 EU-VO zur Anwendung. Hinsichtlich der Anwendung der black list ist somit sowohl die Abgrenzung der Prüferseite als auch die der Leistungsempfänger sehr weit gefasst.⁵¹

Art. 5 Abs. 2 und 3 EU-VO räumt den Mitgliedstaaten die Möglichkeit ein, die Liste der verbotenen Leistungen über die black list hinaus zu erweitern oder auch in gewissem Umfang einzuschränken. So können gem. Art. 5 Abs. 3 EU-VO Bewertungsleistungen sowie die Steuerberatungsleistungen der black list mit Ausnahme der Steuerberatung im Hinblick auf LSt und Zölle insoweit zugelassen werden, als sie allein oder kumuliert keine direkten oder nur unwesentliche Auswirkungen auf die geprüften Abschlüsse haben. Letzteres muss darüber hinaus gegenüber dem Prüfungsausschuss umfassend dokumentiert und erläutert werden und der Abschlussprüfer muss die Grundsätze der Unabhängigkeit gem. der EU-RL erfüllen. Der Prüfungsausschuss erstellt gem. Art. 5 Abs. 4 Unterabs. 1 EU-VO ggf. Leitlinien im Hinblick auf die Erbringung derartiger Leistungen. Auch NPL, die nicht unter das Verbot der black list oder ein vom jeweiligen Mitgliedstaat erweitertes Verbot fallen, können gem. Art. 5 Abs. 4 EU-VO nur mit Billigung des Prüfungsausschusses erbracht werden, sofern die Mitgliedstaaten nicht noch strengere Vorschriften vorsehen.

Die Erbringung von zusätzlichen NPL durch den (Konzern-)Abschlussprüfer war in Deutschland bereits vor Inkrafttreten der EU-VO nur eingeschränkt möglich, wobei die einschlägigen gesetzlichen Vorschriften seit dem Bilanzrechtsreformgesetz (BilReG) für alle (Konzern-)Abschlussprüfer in § 319 HGB a.F. und die zusätzlichen Regelungen für Prüfer von kapitalmarktorientierten Gesellschaften in § 319a HGB a.F. enthalten waren. Seit dem BilMoG regelt § 319b HGB, inwiefern diese Regelungen auch greifen, wenn die NPL jeweils nur von einem Mitglied des Netzwerks des Prüfers erbracht werden. Der in der black list enthaltene Katalog verbotener NPL ist jedoch gegenüber diesem Status quo ante nicht nur hinsichtlich der Art der NPL detaillierter bzw. umfangreicher,⁵² sondern sieht im Gegensatz zur bisherigen Regelung in

⁵¹ Vgl. hierzu ausführlich IDW, Positionspapier zu Nichtprüfungsleistungen des Abschlussprüfers, 4. Aufl. (Stand: 05.11.2018), S. 16 f.

⁵² Vgl. z.B. auch Ratzinger-Sakel/Schönberger, Accounting in Europe 2015 S. 66.

Deutschland auch keine Wesentlichkeitsgrenze vor.⁵³ Insofern erweitert und verschärft die black list der EU-VO die bisher in Deutschland geltenden einschlägigen Regelungen.⁵⁴ Diese Einschätzung dürfte auch dann noch aufrecht erhalten werden können, wenn man berücksichtigt, dass der deutsche Gesetzgeber das oben beschriebene Mitgliedstaatenwahlrecht nach Art. 5 Abs. 3 EU-VO zur Einschränkung der black list hinsichtlich Steuerberatungs- und Bewertungsleistungen im AReG im weitest möglichen Umfang ausgeübt hat.⁵⁵ Dabei hat sich der Gesetzgeber auch bemüht, durch genauere Vorgaben in § 319a Abs. 1 Satz 1 Nr. 2 HGB zu operationalisieren, wann eine unmittelbare Auswirkung einer Steuerberatungsleistung auf den jeweils geprüften Abschluss als nicht nur unwesentlich einzustufen und damit für den (Konzern-)Abschlussprüfer verboten ist. Auch hier dürften sich, wie bei anderen Tatbeständen der black list teilweise auch, Auslegungsspielräume und damit Rechtsunsicherheiten ergeben, so etwa hinsichtlich der entscheidenden Frage, ob bereits eine sog. aggressive Steuerpolitik vorliegt.⁵⁶ Berücksichtigt man die oben erläuterte, bereits in der EU-VO geforderte Billigung der NPL durch den Prüfungsausschuss,⁵⁷ die erst nach gebührender Beurteilung der potenziellen Gefährdung der Unabhängigkeit des (Konzern-)Abschlussprüfers und der verwendeten Schutzmaßnahmen erfolgen kann, sowie die durch das AReG eingeführten einschlägigen Sanktionsvorschriften für den Fall der Zuwiderhandlung,⁵⁸ so könnte dies möglicherweise dazu führen, dass sich Prüfungsausschüsse zukünftig für eine stärkere Trennung von APL und NPL entscheiden könnten als vor Inkrafttreten von EU-VO und AReG.⁵⁹

Die EU-VO geht offensichtlich davon aus, dass eine mit der Abschlussprüfung kombinierte Erbringung von NPL generell eine Gefahr für die Unabhängigkeit des (Konzern-)Abschlussprüfers

⁵³ Vgl. hierzu auch IDW, a.a.O. (Fn. 51), S. 30 i.V.m. S. 29, sowie zur Herausnahme offensichtlicher Bagatellfälle m.w.N. auch Schmidt/Nagel, in: Grottel u.a. (Hrsg.), Beck Bil-Komm., 11. Aufl. 2018, § 319a HGB Rdn. 74.

⁵⁴ Vgl. ebenso Ratzinger-Sakel/Schönberger, Accounting in Europe 2015 S. 66.

⁵⁵ Vgl. § 319a Abs. 1 Satz 1 Nr. 2, 3 i.V.m. Abs. 2 Satz 1 HGB. Vgl. in diesem Zusammenhang z.B. auch IDW, a.a.O. (Fn. 51), S. 32, wo darauf hingewiesen wird, dass im Hinblick auf Lohnsteuern und Zölle Steuerberatungsleistungen im Gegensatz zur bisher in Deutschland geltenden Rechtslage nicht mehr erlaubt sind.

⁵⁶ Vgl. ebenso Velte, WPg 2016 S. 127. Vgl. zur Auslegung z.B. Schmidt/Nagel, a.a.O. (Fn. 53), § 319a HGB Rdn. 15 ff. Vgl. zu aggressiver Steuerpolitik als Hinweis auf Wesentlichkeit insb. Erwägungsgrund 9 der EU-VO.

⁵⁷ Für die im Rahmen des Mitgliedstaatenwahlrechts in Deutschland zugelassenen Steuerberatungsleistungen folgt dieses Erfordernis unmittelbar aus § 319a Abs. 3 HGB, wobei der Gesamtaufsichts- bzw. Verwaltungsrat diese Aufgabe wahrnimmt, falls kein Prüfungsausschuss eingerichtet ist.

⁵⁸ Vgl. hierzu insb. §§ 404a, 405 Abs. 3b AktG.

⁵⁹ Auf durch zusätzliche Sanktionsmechanismen erhöhte Anreize der Prüfungsausschüsse, sich um Einhaltung ihrer Überwachungspflichten gegenüber dem Abschlussprüfer ernsthaft zu bemühen, wird bereits hingewiesen bei Velte, WPg 2016 S. 130.

und damit für die Gewährleistung einer hohen Abschlussprüfungsqualität darstellt, wobei bei zahlreichen Formen von NPL die Gefahr als so hoch eingeschätzt wird, dass deren Kombination mit APL von vornherein verboten wird.⁶⁰ Aus theoretischer Sicht erscheint diese Annahme weniger zwingend, weil in diesem Zusammenhang u.a. auf die Möglichkeit sog. Knowledge-Spillover-Effekte hingewiesen werden kann, die unter bestimmten Bedingungen sogar die Unabhängigkeit des Prüfers und damit die Prüfungsqualität erhöhen könnten.⁶¹ Auch aus empirischer Sicht sind die bisherigen Befunde in dieser Hinsicht nicht eindeutig.⁶²

Vor diesem Hintergrund soll im Folgenden unter Heranziehung der in Abschn. 2.3 beschriebenen Honorarangaben nach § 314 Abs. 1 Nr. 9 HGB für die Stichprobe mit Fehlermeldungen und damit Fällen mit tendenziell niedriger Prüfungsqualität (Fehlergruppe) zunächst festgestellt werden, wie häufig in diesen Fällen im Jahr des Fehlers vom Konzernabschlussprüfer überhaupt zusätzlich zu APL auch NPL erbracht wurden. Darüber hinaus wird für die Fälle, in denen NPL vorliegen, auch ausgewertet, welche der drei neben APL vorgesehenen Kategorien an NPL gem. § 314 Abs. 1 Nr. 9 HGB jeweils erbracht wurden. Da sich die Aufschlüsselung der Honorare auf die zweite und vierte Kategorie, wie in Abschn. 2.3 bereits erläutert, durch das BilMoG geändert hat, wird die Auswertung dabei auch jeweils getrennt für die Fälle vor und nach Inkrafttreten des BilMoG vorgenommen. Zum Vergleich wird in einem zweiten Schritt sodann für die Fälle der Kontrollstichprobe mit tendenziell hoher Prüfungsqualität (Kontrollgruppe) die gleiche Auswertung durchgeführt. Dabei soll zum einen überprüft werden, ob der Anteil der Fälle, bei denen zusätzlich zu APL auch NPL erbracht werden, hier niedriger ist. Zum andern können auch Unterschiede bei der Häufigkeit des Auftretens der einzelnen Kategorien an NPL untersucht werden. Hierbei könnte im vorliegenden Zusammenhang z.B. insb. die dritte Kategorie der StBL von Interesse sein, weil diese einen bedeutenden Teil der black list ausmachen, gleichzeitig von der EU über das eingeräumte Mitgliedstaatenwahlrecht in diesem Zusammenhang jedoch eine erhebliche Abschwächung des strikten Verbots ermöglicht wurde.⁶³ In einem letzten Schritt wird

⁶⁰ Vgl. hierzu auch Erwägungsgrund 8 der EU-VO.

⁶¹ Vgl. ausführlich Beck/Frecka/Solomon, *Journal of Accounting Literature* 1988 insb. S. 51 (58 f.). Auch bei Arruñada, *The Economics of Audit Quality*, 1999, S. 82-85, können unter bestimmten Bedingungen positive Wirkungen des zusätzlichen Angebots von NPL auf die Unabhängigkeit des Abschlussprüfers abgeleitet werden.

⁶² Vgl. zu einer knappen Diskussion aus theoretischer und empirischer Sicht z.B. Wagenhofer/Ewert, a.a.O. (Fn. 6), S. 525-529.

⁶³ Bewertungsleistungen, die ebenfalls vom Mitgliedstaatenwahlrecht erfasst sind, können hingegen mittels der Honorarangaben gem. § 314 Abs. 1 Nr. 9 HGB kaum isoliert werden, weil sie sowohl vor als auch nach dem BilMoG jeweils mit anderen Arten von NPL in einer Kategorie zusammengefasst waren.

schließlich der 1-1-Match explizit berücksichtigt, indem aus allen aus Fehler- und Vergleichsunternehmen gebildeten Paaren jeweils diejenigen herausgefiltert werden, die hinsichtlich des Auftretens der jeweils betrachteten Honorarkategorie nicht übereinstimmen (Bestimmung der sog. discordant pairs). Anhand einer einfachen univariaten Analyse kann so für diese Paare jeweils bestimmt werden, wie sich die Wahrscheinlichkeit dafür, dass das Unternehmen, bei dem Honorare der jeweiligen Kategorie anfallen, ein Unternehmen mit einer Fehlermeldung ist, zu der Gegenwahrscheinlichkeit verhält, dass es sich um kein Unternehmen mit einer Fehlermeldung handelt.⁶⁴

Zu berücksichtigen ist allerdings, dass diese Auswertung verschiedenen Einschränkungen unterliegt. So wird in der neuen Kommentarliteratur z.B. darauf hingewiesen,⁶⁵ dass die Definition von APL gem. § 314 Abs. 1 Nr. 9 HGB tendenziell etwas umfassender sein dürfte als die einschlägige Definition des Art. 2 Nr. 1 EU-RL, sodass nicht für jeden Fall im Folgenden sicher gestellt werden kann, dass die erste Kategorie der ausgewerteten Honorarangaben nicht auch Honorare für Leistungen enthält, die aus EU-Sicht bereits NPL i.S.d. Art. 5 EU-VO darstellen würden. Im Übrigen wurde bereits in Abschn. 2.3 darauf hingewiesen, dass bei der Veröffentlichung der ausgewerteten Honorarangaben hinsichtlich der inhaltlichen Zuordnung zu den Kategorien aufseiten des verantwortlichen Managements ein gewisser Gestaltungsspielraum besteht.⁶⁶ Auch ist die Abgrenzung des Abschlussprüfers für die Zwecke der zur Verfügung stehenden Honorarangaben selbst vor Inkrafttreten des BilMoG enger als in Art. 5 EU-VO. Deshalb wird für die Auswertung in Tab. 2.3 bei den wenigen Fällen, bei denen die Honorarangaben alternativ für unterschiedliche Abgrenzungen auf Prüferseite angegeben sind, jeweils von der weitesten Abgrenzung ausgegangen, sofern nicht ohnehin unabhängig von der Abgrenzung jeweils die gleichen Kategorien an Leistungen angeboten wurden.⁶⁷ Was die

⁶⁴ Vgl. hierzu Hosmer/Lemeshow/Sturdivant, Applied Logistic Regression, 3. Aufl. 2013, S. 243-246. Zurückgegriffen wird hierbei auf eine univariate, sog. bedingte logistische Regression.

⁶⁵ Vgl. Schmidt/Nagel, a.a.O. (Fn. 53), § 319a HGB Rdn. 71.

⁶⁶ Im Übrigen werden die im Folgenden auszuwertenden Angaben auch nicht in jedem Fall entsprechend der vier vorgegebenen Kategorien gemacht. Bei diesen Fällen werden die Honorarangaben, wenn die Zuordnung nicht ohnehin eindeutig ist, für die folgende Auswertung der plausibelsten Kategorie zugeordnet. So werden der Kategorie sL z.B. grds. auch Honorare für „sonstige Beratung“ oder „prüfungsnaher Beratung bzw. Dienstleistungen“ zugeordnet; vgl. zu Letzterem insb. Ellrott, a.a.O. (Fn. 37), § 285 HGB Rdn. 274, sowie Begr RegE BilMoG BT-Drucks. 16/10067, S. 71. Dabei gestaltet sich die Abgrenzung zwischen der Kategorie aBL und sL besonders schwierig.

⁶⁷ Alternative Abgrenzungen auf Prüferseite sind bei sieben (sechs) der ausgewerteten Fälle der Fehlergruppe (Kontrollgruppe) vorhanden, wobei sich davon lediglich bei jeweils zwei Fällen der Fehlergruppe bzw. der Kontrollgruppe die Kategorien der angebotenen Leistungen unterscheiden.

Abgrenzung der Leistungsempfänger im Konzern betrifft,⁶⁸ so ist im Fall eines Teilkonzernabschlusses die Abgrenzung nach § 314 Abs. 1 Nr. 9 HGB über den gesamten Zeitraum insofern enger als nach Art. 5 EU-VO,⁶⁹ als Art. 5 EU-VO zumindest auch die Honorare für Leistungen an das übergeordnete Mutterunternehmen miteinschließt. Für die Zeit nach dem BilMoG ist hingegen zu beachten, dass Honorare für Leistungen an Gemeinschaftsunternehmen in Deutschland quotal miteinbezogen werden sollen, nach Art. 5 Abs. 1 EU-VO hingegen nicht. Schließlich ist zu berücksichtigen, dass dasselbe Fehler- und dasselbe Vergleichsunternehmen für unterschiedliche Jahre jeweils mehrfach in der Fehler- bzw. Kontrollgruppe enthalten sein können.⁷⁰ Die beobachteten NPL-Kategorien sind für diese Fälle möglicherweise nicht völlig unabhängig voneinander.

Die Auswertung der im Konzernanhang jeweils angegebenen Honorarkategorien ist in Tab. 2.3 dargestellt.⁷¹ Auffällig ist zunächst, dass in der Fehlergruppe bei 21,7% der Beobachtungen und damit bei mehr als jedem fünften Fall mit fehlerhaftem Konzernabschluss bzw. -lagebericht und damit tendenziell niedriger Prüfungsqualität im Jahr des Fehlers überhaupt keine NPL vom Konzernabschlussprüfer bezogen werden. Demgegenüber trifft dies in der Kontrollgruppe mit tendenziell hoher Prüfungsqualität im jeweiligen Jahr nur bei 11,9% zu, sodass hier nur in etwa jedem neunten Fall keine NPL erbracht werden. Zwar ist dieser Anteil in der Kontrollgruppe nach Inkrafttreten des BilMoG und damit in der jüngeren Vergangenheit deutlich höher, trotzdem liegt der Anteil der Fälle ohne NPL hier immer noch um 5,6 Prozentpunkte unterhalb des entsprechenden Anteils in der Fehlergruppe.

⁶⁸ Vgl. auch hierzu hinsichtlich § 314 Abs. 1 Nr. 9 HGB die Ausführungen in Abschn. 2.3.

⁶⁹ Dies gilt zumindest, wenn man Art. 5 Abs. 1 und Art. 5 Abs. 5 EU-VO gemeinsam betrachtet, auch wenn Art. 5 Abs. 5 EU-VO für NPL bei Tochterunternehmen der PIE mit Sitz außerhalb der EU etwas weniger restriktive Vorgaben macht.

⁷⁰ Zur Begründung dieser Vorgehensweise vgl. Abschn. 2.3.

⁷¹ Acht der 151 in Tab. 2.2 abgegrenzten Fälle der Stichprobe der Fehlerveröffentlichungen können nicht einbezogen werden. Zusätzlich zu den in Abschn. 2.3 bereits erläuterten jeweils sieben Fällen in Fehler- und Kontrollgruppe, die aufgrund der im jeweiligen Geschäftsjahr noch nicht bestehenden Honorarangabepflicht nicht ausgewertet werden können, existiert in der Stichprobe der Fehlermeldungen ein Fall, bei dem die Honorardaten trotz bereits bestehender Veröffentlichungsverpflichtung nicht im Konzernanhang angegeben sind. Aufgrund des durchgeführten 1-1-Match wird damit nicht nur dieser Fall, sondern auch die im Rahmen des 1-1-Match zugeordnete Unternehmens-Jahr-Beobachtung aus der Kontrollstichprobe ausgeschlossen.

Tab. 2.3: Häufigkeit des Auftretens von Abschlussprüfungs- und Nichtprüfungsleistungen im Jahr des Fehlers

Fehlergruppe	Abschluss- Prüfungs- leistungen^{a)}	Davon ausschließlich Abschluss-prüfungs- leistungen	Andere Bestätigungs- leistungen^{b)}	Steuer- beratungs- leistungen	Sonstige Leistungen
Häufigkeit des Auftretens der Kategorie	143 (100%)	31 (21,7%)	64 (44,8%)	59 (41,3%)	79 (55,2%)
<i>vor BilMoG</i>	75 (100%)	15 (20,0%)	33 (44,0%)	33 (44,0%)	43 (57,3%)
<i>nach BilMoG</i>	68 (100%)	16 (23,5%)	31 (45,6%)	26 (38,2%)	36 (52,9%)
Stichprobenumfang zur Ermittlung der Honorarkategorien	143	143	143	143	143
Kontrollgruppe	Abschluss- prüfungs- leistungen	Davon ausschließlich Abschluss-prüfungs- leistungen	Andere Bestätigungs- leistungen	Steuer- beratungs- leistungen	Sonstige Leistungen
Häufigkeit des Auftretens der Kategorie	143 (100%)	17 (11,9%)	63 (44,1%)	91 (63,6%)	102 (71,3%)
<i>vor BilMoG</i>	76 (100%)	5 (6,6%)	36 (47,4%)	51 (67,1%)	55 (72,4%)
<i>nach BilMoG</i>	67 (100%)	12 (17,9%)	27 (40,3%)	40 (59,7%)	47 (70,1%)
Stichprobenumfang zur Ermittlung der Honorarkategorien	143	143	143	143	143

^{a)} Dass sich die Fallzahl vor und nach Inkrafttreten des BilMoG in der Fehler- und Kontrollgruppe jeweils um einen Fall unterscheidet, ist auf vom Kalenderjahr abweichende Geschäftsjahre zurückzuführen.

^{b)} Vor dem Inkrafttreten des BilMoG umfasst diese Kategorie die sonstigen Bestätigungs- oder Bewertungsleistungen. Nach dem Inkrafttreten des BilMoG sind die Bewertungsleistungen im Auffangtatbestand sL auszuweisen. Wird nach Inkrafttreten des BilMoG fälschlicherweise noch die Bezeichnung „sonstige Bestätigungs- oder Bewertungsleistungen“ verwendet, so werden im Folgenden die Honorare trotzdem in der Kategorie aBL erfasst.

Hinsichtlich der zweiten Honorarkategorie gibt es über den Gesamtzeitraum keinen wesentlichen Unterschied zwischen den beiden Gruppen. In 44,8% (44,1%) der Fälle in der Fehlergruppe (Kontrollgruppe) werden NPL dieser Kategorie erbracht. Allerdings ist zu beachten, dass diese Kategorie erst nach Inkrafttreten des BilMoG ausschließlich Honorare für aBL enthält, die in der EU-VO teilweise explizit von der black list ausgenommen sind⁷² und damit aufgrund ihrer Natur offensichtlich als weniger bedenklich für die Unabhängigkeit des Prüfers angesehen werden.⁷³ In diesem Zusammenhang ist das nicht tabellierte Ergebnis von Bedeutung, dass in der Fehlergruppe bei 11,8% aller 68 Beobachtungen nach Inkrafttreten des BilMoG ausschließlich APL sowie aBL erbracht werden. Addiert man hierzu die Fälle, in denen ausschließlich APL bezogen werden, so ergibt sich nach Inkrafttreten des BilMoG ein Anteil von immerhin 35,3% der Fälle der

⁷² Vgl. hierzu Art. 5 Abs. 1 Unterabs. 2i) EU-VO, wo Bestätigungsleistungen im Zusammenhang mit den Abschlüssen explizit von der black list ausgenommen werden, wobei comfort letters im Hinblick auf vom geprüften Unternehmen herausgegebene Prospekte als Beispiel genannt werden.

⁷³ Vgl. hierzu auch Erwägungsgrund 8 der EU-VO.

Fehlergruppe, bei denen eher keine aus EU-Sicht als problematisch angesehenen NPL bezogen werden. Während dies für etwa jeden dritten Fall der Fehlergruppe gilt, trifft es hingegen nur auf etwas mehr als jeden sechsten Fall der Kontrollgruppe zu, da der entsprechende, ebenfalls nicht tabellierte Anteil für die Fälle nach Inkrafttreten des BilMoG hier nur 17,9% beträgt.⁷⁴ Was schließlich die StBL und die Auffangkategorie der sL betrifft, so werden diese in 41,3% bzw. 55,2% aller Beobachtungen der Fehlergruppe erbracht, wohingegen in beiden Kategorien der Anteil in der Kontrollgruppe mit tendenziell hoher Prüfungsqualität deutlich höher, nämlich bei 63,6% bzw. 71,3% liegt. Insofern sprechen diese deskriptiven Daten im Hinblick auf die StBL zumindest nicht gegen die Abschwächung des strikten Verbots, das in dem von der EU in diesem Zusammenhang eingeräumten Mitgliedstaatenwahlrecht zum Ausdruck kommt.⁷⁵

Allerdings ist zu berücksichtigen, dass Fehler- und Kontrollgruppe bisher lediglich separat voneinander ausgewertet wurden, ohne die 1-1-Zuordnung des Matchingprozesses explizit in die Analyse miteinzubeziehen.⁷⁶ Diese soll im vorliegenden Zusammenhang jedoch gerade erlauben, bei jeweils gleicher Ausprägung der Matchingkriterien⁷⁷ den Zusammenhang zwischen dem Auftreten eines Fehlers und dem Auftreten der jeweiligen Honorarkategorie zu analysieren. Hierzu werden zunächst für jede der vier Honorarkategorien aus allen aus Fehler- und Vergleichsunternehmen gebildeten Paaren jeweils diejenigen herausgefiltert, die hinsichtlich des

⁷⁴ Der Anteil entspricht dem tabellierten Anteil der Fälle, bei denen nur APL ausgewiesen werden, weil es nach Inkrafttreten des BilMoG in der Kontrollgruppe keinen Fall gibt, bei dem ausschließlich APL und aBL auftreten.

⁷⁵ Vgl. in diesem Zusammenhang auch die in die gleiche Richtung weisende Schlussfolgerung bei Hohenfels/Quick, *Review of Managerial Science OnlineFirst* 25.10.2018.

⁷⁶ Bei Cram/Karan/Stuart, *Contemporary Accounting Research* 2009 insb. S. 479 ff., wird auf die Notwendigkeit hingewiesen, für eine korrekte statistische Analyse mit diskreter abhängiger Variable das Matching mittels einer bedingten logistischen Regression explizit zu berücksichtigen. Vgl. zur methodischen Vorgehensweise bei der univariaten Analyse im Folgenden auch Hosmer/Lemeshow/Sturdivant, a.a.O, (Fn. 64), Kap. 7.

⁷⁷ Eine Ausnahme bildet insoweit das Kriterium der Unternehmensgröße, weil hier für die Matchingpartner keine exakte Größengleichheit gewährleistet ist. Die in Tab. 2.4 präsentierte univariate Analyse kontrolliert insofern nicht perfekt für den Einfluss der Unternehmensgröße; vgl. hierzu insb. Cram/Karan/Stuart, *Contemporary Accounting Research* 2009 S. 480. Eine durchgeführte Robustheitsanalyse (vgl. hierzu Cram/Karan/Stuart, *Contemporary Accounting Research* 2009 S. 494-497), bei der alternativ die Unternehmensgröße bzw. die logarithmierte Unternehmensgröße zusätzlich als unabhängige Variable bei den bedingten logistischen Regressionen berücksichtigt wird, führt jedoch hinsichtlich Vorzeichen des in Tab. 2.4, Spalte 7, präsentierten Koeffizienten und Größenordnung des p-Werts nahezu durchgängig zu keinem abweichenden Ergebnis. Lediglich in der letzten Zeile ist der Koeffizient von D nur noch schwach signifikant ($p = 0,059$), wenn man die nicht logarithmierte Unternehmensgröße verwendet. Fragt man für diese Zeile hingegen, wie das Auftreten von sL nach BilMoG vom Auftreten einer Fehlermeldung (Zugehörigkeit zur Fehlergruppe) und der nicht logarithmierten Unternehmensgröße abhängt, so ist der Koeffizient des Auftretens einer Fehlermeldung nicht mehr signifikant ($p = 0,115$).

Auftretens der betrachteten Honorarkategorie nicht übereinstimmen.⁷⁸ Deren Anzahl ist Tab. 2.4 zu entnehmen, wobei danach unterschieden wird, ob die betrachtete Kategorie beim Fehlerunternehmen (Spalte 2) oder beim Vergleichsunternehmen (Spalte 3) des jeweils betrachteten 1-1-Match angegeben ist. Für die Kategorie der APL und StBL erfolgt die Auswertung dabei ausschließlich für den Gesamtzeitraum und damit für alle 143 Paare, während für die aBL und sL zusätzlich auch eine Auswertung nur für die Zeit nach Inkrafttreten des BilMoG erfolgt.⁷⁹

Das Verhältnis der in Spalte 2 von Tab. 2.4 eingetragenen Anzahl zur Anzahl von Paaren in Spalte 3 entspricht hinsichtlich der jeweils betrachteten Honorarkategorie der sog. odds ratio, die jeweils Spalte 4 zu entnehmen ist. Der Quotient aus der jeweiligen Anzahl in Spalte 2 dividiert durch die Summe der für die jeweilige Kategorie insgesamt auftretenden nicht übereinstimmenden Paare (Summe aus Spalte 2 und 3) führt zu der in Spalte 5 angegebenen Wahrscheinlichkeit.⁸⁰ Die Gegenwahrscheinlichkeit hierzu findet sich in Spalte 6. Spalte 7 enthält schließlich den mittels einer univariaten bedingten logistischen Regression geschätzten Koeffizienten,⁸¹ wenn als unabhängige Variable das Auftreten der Honorarkategorie der jeweiligen Zeile und als abhängige Variable das Auftreten einer Fehlermeldung verwendet wird.⁸²

Die Ergebnisse der paarweisen Analyse in Tab. 2.4 lassen die in Tab. 2.3 präsentierten Unterschiede bei den Honorarkategorien eher noch stärker hervortreten. So beziehen z.B. bei 25 der 143 Paare (1-1-Matches) lediglich die Unternehmen mit Fehlermeldung ausschließlich APL vom Konzernabschlussprüfer, wohingegen dies nur bei 11 Paaren ausschließlich auf das jeweilige Vergleichsunternehmen aus der Kontrollgruppe zutrifft. Die Wahrscheinlichkeit dafür, dass bei einem nicht übereinstimmenden Paar ein Unternehmen, das ausschließlich APL bezieht, ein Unternehmen mit Fehlermeldung ist, beträgt somit 69,44% und ist damit 2,27 mal so groß wie die

⁷⁸ Hinsichtlich des Auftretens bzw. Nichtauftretens der jeweiligen Kategorie übereinstimmenden Paaren (sog. concordant pairs) kann hingegen kein zusätzlicher Erklärungsgehalt zukommen; vgl. hierzu Hosmer/Lemeshow/Sturdivant, a.a.O. (Fn. 64), S. 246.

⁷⁹ Dass hierbei nur 66 auszuwertende Paare verbleiben, sich gegenüber Tab. 2.3 also die untersuchten Fälle nach Inkrafttreten des BilMoG in der Fehlergruppe von 68 auf 66 und in der Kontrollgruppe von 67 auf 66 verringern, ist auf Paare zurückzuführen, bei denen jeweils nur beim Fehler- oder beim Kontrollunternehmen ein vom Kalenderjahr abweichendes Geschäftsjahr 2008/09 vorliegt.

⁸⁰ Diese entspricht dem Quotienten $\text{odds ratio}/(1+\text{odds ratio})$.

⁸¹ Dieser entspricht jeweils dem natürlichen Logarithmus der odds ratio. Der p-Wert in Klammern gibt bei Durchführung eines zweiseitigen Tests an, ob der Koeffizient mit 10% (*), 5% (**) oder 1% (***) Irrtumswahrscheinlichkeit signifikant von Null verschieden ist.

⁸² Beide jeweils definiert als dichotome Variable mit Ausprägung Eins im Fall des Auftretens und Null sonst.

Wahrscheinlichkeit, dass es sich bei diesem Unternehmen um ein Unternehmen ohne Fehlermeldung handelt. Analog können auch die weiteren Zeilen von Tab. 2.4 interpretiert werden.

Tab. 2.4: Paarweiser Vergleich von Abschlussprüfungs- und Nichtprüfungsleistungen im Jahr des Fehlers

Dichotome Variable D	Anzahl an Paaren mit D = 1 beim Fehlerunternehmen und D = 0 beim Vergleichsunternehmen	Anzahl an Paaren mit D = 0 beim Fehlerunternehmen und D = 1 beim Vergleichsunternehmen	Odds Ratio	Wahrscheinlichkeit, dass bei einem nicht übereinstimmenden Paar ein Unternehmen mit D = 1 das Fehlerunternehmen ist.	Wahrscheinlichkeit, dass bei einem nicht übereinstimmenden Paar ein Unternehmen mit D = 0 das Fehlerunternehmen ist.	Koeffizient der unabhängigen Variable D einer univariaten bedingten logistischen Regression (p-Wert, zweiseitig) ^{a)}
Nur APL = 1; nicht nur APL = 0 (n = 143 Paare)	25	11	2,27	69,44%	30,56%	0,821 (0,023**)
weder StBL noch sL = 1; StBL und/oder sL = 0 (nach BilMoG, n = 66 Paare)	17	5	3,40	77,27%	22,73%	1,224 (0,016**)
aBL = 1; keine aBL = 0 (n = 143 Paare)	30	29	1,03	50,85%	49,15%	0,034 (0,896)
aBL = 1; keine aBL = 0 (nach BilMoG, n = 66 Paare)	13	10	1,30	56,52%	43,48%	0,262 (0,533)
StBL = 1; keine StBL = 0 (n = 143 Paare)	21	53	0,40	28,38%	71,62%	-0,926 (0,000***)
sL = 1; keine sL = 0 (n = 143 Paare)	20	43	0,47	31,75%	68,25%	-0,765 (0,005***)
sL = 1; keine sL = 0 (nach BilMoG, n = 66 Paare)	7	18	0,39	28,00%	72,00%	-0,944 (0,034**)

^{a)} Koeffizient der unabhängigen Variable und p-Wert der univariaten bedingten logistischen Regression bleiben im Übrigen generell unverändert, wenn D als abhängige und das Auftreten einer Fehlermeldung (Zugehörigkeit zur Fehlergruppe) als unabhängige Variable verwendet wird.

Zusammenfassend kann konstatiert werden, dass es sich bei den Fällen, bei denen nur APL oder nach Inkrafttreten des BilMoG ausschließlich APL und aBL erbracht werden, eher um Fälle der Fehlergruppe und damit Fälle mit tendenziell niedriger Prüfungsqualität handelt. StBL und sL treten hingegen deutlich häufiger in der Kontrollgruppe und somit bei Fällen mit tendenziell hoher Prüfungsqualität auf. Bei ausschließlicher Betrachtung der Häufigkeit des Auftretens von Prüfungs- und Nichtprüfungsleistungen der einzelnen Kategorien ergeben sich also offensichtlich

keine Hinweise dafür, dass den von der EU als kritisch angesehenen NPL bei den Fällen der Fehlergruppe eine höhere Bedeutung zukommt als in der Kontrollgruppe. Ein Argument für die weitere Verschärfung der Regelungen hinsichtlich des Angebots zusätzlicher NPL durch Einführung der black list ergibt sich hieraus also nicht.⁸³ Die Ergebnisse der Analyse hinsichtlich der StBL sprechen auch nicht gegen die Abschwächung des strikten Verbots, das in dem von der EU in diesem Zusammenhang eingeräumten Mitgliedstaatenwahlrecht zum Ausdruck kommt. Bisher wurde allerdings nur die Art der vom jeweiligen Konzernabschlussprüfer erbrachten Leistungen betrachtet. Gegenstand des zweiten Teils des Beitrags werden hingegen die Begrenzungen des Umfangs sein, die die EU-VO für die Honorare aus NPL bzw. die Gesamthonorare vorsieht.

2.5 Zusammenfassung zu Teil 1 und Ausblick

Mit dem Ziel der Verbesserung der Qualität der Abschlussprüfung hat die Verordnung (EU) Nr. 537/2014 über spezifische Anforderungen an die Abschlussprüfung bei Unternehmen von öffentlichem Interesse vom April 2014 u.a. die Erbringung von NPL durch den Abschlussprüfer sowohl hinsichtlich der Kategorien zulässiger NPL als auch im Hinblick auf den erlaubten Umfang weiter eingeschränkt. Auch mit der erstmaligen verbindlichen Einführung der sog. externen Rotation soll die Prüfungsqualität verbessert werden.

In Teil 1, Teil 2 und Teil 3 dieser Studie werden anhand einer deskriptiven Analyse Hinweise zur Beantwortung der Frage geliefert, ob diese Vorgaben der EU-VO sowie die damit zusammenhängende Ausübung bzw. Nichtausübung in der EU-VO vorgesehener Mitgliedstaatenwahlrechte im Rahmen des AReG tatsächlich geeignet sind, zur Verbesserung der Abschlussprüfungsqualität beizutragen. Hierzu wird zunächst in Teil 1 des Beitrags für eine Stichprobe von Fällen mit tendenziell niedriger Prüfungsqualität untersucht, inwieweit die durch die EU-VO vorgesehenen Beschränkungen hinsichtlich der Art von NPL bei diesen Fällen überhaupt zum Tragen gekommen wären, wenn sie im Untersuchungszeitraum bereits zu beachten gewesen wären. Als Surrogatmaß für eine niedrige Prüfungsqualität werden dabei zwischen dem 01.07.2005 und dem 31.12.2017 im Bundesanzeiger veröffentlichte Meldungen von Rechnungslegungsfehlern verwendet, die von der DPR bzw. der BaFin im Rahmen eines sog.

⁸³ Auf die o.g. Einschränkungen der Auswertung sei an dieser Stelle noch einmal hingewiesen.

Enforcementverfahren festgestellt werden. Die Bestandsaufnahme wird jeweils mit dem Befund einer Kontrollstichprobe mit tendenziell hoher Prüfungsqualität verglichen.

Bei ausschließlicher Betrachtung der Häufigkeit des Auftretens der im Anhang des Konzernabschlusses gem. § 314 Abs. 1 Nr. 9 HGB zu veröffentlichenden Kategorien von Prüfungs- und Nichtprüfungsleistungen ergeben sich keine Hinweise dafür, dass den von der EU als kritisch angesehenen NPL bei den Fällen der Fehlergruppe eine höhere Bedeutung zukommt als in der Kontrollgruppe. Vielmehr sind Unternehmen, bei denen nur APL oder nach Inkrafttreten des BilMoG ausschließlich APL und aBL erbracht werden, eher Unternehmen mit Fehlermeldungen. StBL und sL sind hingegen deutlich häufiger in der Kontrollgruppe vertreten. Insofern ergibt sich hieraus kein Argument für die weitere Verschärfung der bisherigen Regelungen hinsichtlich des Angebots zusätzlicher NPL durch Einführung der black list. Auch sprechen die Ergebnisse der Analyse im Hinblick auf die StBL zumindest nicht gegen die Abschwächung des strikten Verbots, das in dem von der EU in diesem Zusammenhang eingeräumten Mitgliedstaatenwahlrecht zum Ausdruck kommt.⁸⁴ In Teil 2 dieses Beitrags wird die empirische Auswertung auf den Umfang angebotener Prüfungs- und Nichtprüfungsleistungen in der Fehler- und Kontrollgruppe ausgedehnt.

2.6 Einleitung zu Teil 2

In der Literatur finden sich durchaus Zweifel, ob die konkreten Vorgaben der Verordnung (EU) Nr. 537/2014 (EU-VO) über spezifische Anforderungen an die Abschlussprüfung bei Unternehmen von öffentlichem Interesse (PIEs) vom April 2014 tatsächlich geeignet sind, zur Verbesserung der Abschlussprüfungsqualität beizutragen.⁸⁵ Vor diesem Hintergrund erfolgt in dieser Studie für eine Stichprobe von Fällen mit tendenziell niedriger Prüfungsqualität eine deskriptive Auswertung dahingehend, inwieweit die durch die EU-VO vorgesehenen Beschränkungen hinsichtlich Art und/oder Umfang von Nichtprüfungsleistungen (NPL) bei diesen Fällen überhaupt zum Tragen gekommen wären, wenn sie im Untersuchungszeitraum bereits zu beachten gewesen wären. Die Ausübung bzw. Nichtausübung in der EU-VO vorgesehener Mitgliedstaatenwahlrechte im Rahmen des am 17.06.2016 in Kraft getretenen

⁸⁴ Wie jede empirische Untersuchung unterliegt auch die vorliegende Studie spezifischen Einschränkungen. Diese werden in der Schlussbetrachtung des Gesamtbeitrags in Teil 3, Abschn. 2.11, zusammengefasst.

⁸⁵ Vgl. im Hinblick auf black list und cap z.B. Ratzinger-Sakel/Schönberger, *Accounting in Europe* 2015 S. 62 f.; im Hinblick auf die externe Rotation vgl. z.B. Pott/Schröder/Weckelmann, *KoR* 2014 S. 504.

Abschlussprüfungsreformgesetzes (AReG) wird dabei jeweils mit berücksichtigt. Betrachtet wird schließlich auch, inwieweit bei diesen Fällen mit tendenziell niedriger Prüfungsqualität eine freiwillige Prüferrotation zu beobachten war. Die Bestandsaufnahme wird jeweils mit dem Befund einer Kontrollstichprobe verglichen.

Da als Surrogatmaß für eine niedrige Prüfungsqualität in dieser Untersuchung zwischen dem 01.07.2005 und dem 31.12.2017 im Bundesanzeiger veröffentlichte Meldungen von Rechnungslegungsfehlern verwendet werden, die von der Deutschen Prüfstelle für Rechnungslegung (DPR) bzw. der BaFin im Rahmen eines sog. Enforcementverfahrens festgestellt wurden, wurde in Teil 1 dieses Beitrags (Abschn. 2.1-2.5) zunächst der Begriff der Prüfungsqualität erläutert und verdeutlicht, wo die Grenzen des gewählten Surrogatmaßes liegen. Darauf folgte zum einen die Bereinigung und genaue Abgrenzung der Stichprobe der Fehlermeldungen (Fehlergruppe). Zum anderen wurden zu der verbleibenden Stichprobe von 151 Fehlermeldungen mittels eines sog. 1-1-Match eine Kontrollstichprobe (Kontrollgruppe) abgegrenzt⁸⁶ und die weiteren für die Untersuchung erhobenen Daten erläutert. Schließlich wurde eine erste deskriptive Analyse der in Fehler- und Kontrollgruppe auftretenden Kategorien von Prüfungs- und Nichtprüfungsleistungen vorgenommen. Hieraus konnte unter dem Gesichtspunkt der Verbesserung der Prüfungsqualität kein Argument für die Verschärfung des Verbots von NPL durch die EU-VO abgeleitet werden. Der zweite Teil dieses Beitrags widmet sich nun den in quantitativer Hinsicht von der EU-VO vorgesehenen Einschränkungen bei der Erbringung von Prüfungs- bzw. Nichtprüfungsleistungen. Abschn. 2.7.1 befasst sich dabei zunächst mit der Deckelung der Honorare, die der jeweilige (Konzern-)Abschlussprüfer maximal aus NPL erzielen darf (sog. cap). Abschn. 2.7.2 widmet sich der Begrenzung des maximal zulässigen Anteils der Honorare aus Prüfung und NPL aus einem bestimmten PIE-Mandat am gesamten Umsatz des jeweiligen (Konzern-)Abschlussprüfers aus allen Mandaten. Abschn. 2.8 fasst die wesentlichen Befunde des zweiten Teils der empirischen Bestandsaufnahme zusammen. Vor dem Hintergrund der durch die EU-VO zusätzlich eingeführten externen Prüferrotation ist der demnächst folgende Teil 3 schließlich der Auswertung der freiwilligen Prüferwechsel gewidmet.

⁸⁶ Ein 1-1-Match findet sich in mehreren bisherigen Studien zum deutschen Enforcement; vgl. z.B. Strohmenger, *Journal of International Financial Management & Accounting* 2014 S. 281, sowie Ebner/Hottmann/Zülch, *Corporate Ownership & Control* 2017 S. 128.

2.7 Untersuchungsergebnisse – Obergrenzen für die Honorare aus Nichtprüfungsleistungen und für die Gesamthonorare

2.7.1 Ist die Einführung des cap eine wirksame Maßnahme zur Steigerung der Prüfungsqualität?

Die EU-VO verbietet einem Abschlussprüfer nicht nur die Erbringung bestimmter NPL bei seinem Prüfungsmandanten, sondern legt in Art. 4 Abs. 2 EU-VO auch noch eine maximal zulässige Honorarobergrenze (cap) für nicht in der black list gem. Art. 5 Abs. 1 EU-VO enthaltene NPL fest. Erbringt der Abschlussprüfer für drei oder mehr aufeinanderfolgende Geschäftsjahre für das geprüfte Unternehmen, dessen Muttergesellschaft oder dessen Tochterunternehmen derartige NPL, so dürfen die Honorare für diese NPL ab dem vierten Jahr jeweils 70% der durchschnittlichen Prüfungshonorare der letzten drei Jahre nicht überschreiten.⁸⁷ Die Prüfungshonorare umfassen dabei die Honorare für die Prüfung der Abschlüsse des geprüften Unternehmens und ggf. von dessen Muttergesellschaft und den Tochterunternehmen sowie auch die Honorare für die Prüfung der Konzernabschlüsse der betreffenden Unternehmensgruppe. Aufseiten des (Konzern-)Abschlussprüfers sind hingegen weder Honorare an seine verbundenen Unternehmen noch an Mitglieder seines Netzwerks mit einzurechnen. Damit die Vorschrift greift, müssen nicht nur die NPL, sondern auch die Abschlussprüfungsleistungen (APL) durchgängig in jedem der vergangenen drei Geschäftsjahre vom betrachteten Prüfer erbracht worden sein. Außerdem muss der Prüfungsmandant durchgängig PIE sein.⁸⁸ Nach dem in Art. 4 Abs. 2 EU-VO vorgesehenen Mitgliedstaatenwahlrecht könnte der Abschlussprüferaufsichtsbehörde gestattet werden, dem Prüfer auf dessen Antrag hin im Ausnahmefall zu erlauben, im Hinblick auf ein bestimmtes geprüftes Unternehmen den cap von 70% für maximal zwei Jahre zu überschreiten. Dieses Mitgliedstaatenwahlrecht wurde durch den deutschen Gesetzgeber im AReG nur teilweise ausgeübt. Gem. § 319a Abs. 1a HGB darf die 70%-Grenze mit Ausnahmegenehmigung der Abschlussprüferaufsichtsstelle (APAS) nur maximal für ein Geschäftsjahr überschritten werden, wobei dann ein cap von 140% gilt. Nicht ausgeübt wurde vom deutschen Gesetzgeber im AReG

⁸⁷ Für die Gültigkeit des cap erst ab dem vierten Jahr vgl. insb. EU-Kommission, Q&A – Implementation of the New Statutory Audit Framework, 03.09.2014, S. 2 f.

⁸⁸ Vgl. EU-Kommission, a.a.O. (Fn. 87), S. 3, sowie IDW, Positionspapier zu Nichtprüfungsleistungen des Abschlussprüfers, 4. Aufl. (Stand: 05.11.2018), S. 44 f.

hingegen das Mitgliedstaatenwahlrecht nach Art. 4 Abs. 4 EU-VO, das jedem Mitgliedstaat erlauben würde, strengere Anforderungen, als die in Art. 4 Abs. 2 EU-VO enthaltenen, vorzusehen.

Die Vorschrift zum cap bindet also die Höhe der maximal zulässigen Honorare aus NPL an die Höhe der erzielten Honorare aus APL. Die EU-VO geht nämlich davon aus, dass sowohl die Höhe als auch die Zusammensetzung der Honorare des Abschlussprüfers dessen Unabhängigkeit gefährden könnten.⁸⁹ In der Literatur wird in diesem Zusammenhang nun jedoch zu Recht darauf hingewiesen, dass die 70%-Grenze willkürlich gewählt erscheint und empirische Befunde dafür fehlen, dass durch die Einhaltung des cap tatsächlich die Prüfungsqualität gesteigert werden kann.⁹⁰ Vor diesem Hintergrund soll im Folgenden unter Heranziehung der in Teil 1, Abschn. 2.3, beschriebenen Honorarangaben nach § 314 Abs. 1 Nr. 9 HGB für die Stichprobe mit Fehlermeldungen und damit Fällen mit tendenziell niedriger Prüfungsqualität festgestellt werden, wie häufig der cap im Jahr des Fehlers überhaupt überschritten worden wäre und damit ein nach der EU-VO nicht mehr zulässiges (und somit als qualitätsmindernd eingestuftes) Verhältnis von Nichtprüfungs- zu Abschlussprüfungshonoraren vorgelegen hätte. Zum Vergleich wird in einem zweiten Schritt dann für die Fälle der Kontrollstichprobe mit tendenziell hoher Prüfungsqualität die gleiche Auswertung durchgeführt, um zu überprüfen, ob der Anteil der Fälle, bei denen der cap überschritten worden wäre, hier niedriger ist. Schließlich wird analog zur Vorgehensweise in Teil 1, Abschn. 2.4, der 1-1-Match explizit berücksichtigt, indem aus allen aus Fehler- und Vergleichsunternehmen gebildeten Paaren diejenigen herausgefiltert werden, die hinsichtlich Überschreitung bzw. Nichtüberschreitung des cap nicht übereinstimmen.

Zu berücksichtigen ist allerdings, dass die Berechnung des cap gem. EU-VO mit der beschriebenen Vorgehensweise nur angenähert werden kann. So liefern die zur Verfügung stehenden Honorarangaben nicht exakt die Größen, die nach Art. 4 Abs. 2 EU-VO erforderlich wären. Hinsichtlich der Abgrenzung des Prüfers gilt dies auf jeden Fall für die Honorarangaben vor Inkrafttreten des Bilanzrechtsmodernisierungsgesetzes (BilMoG), weil die Honorare hier i.d.R. nicht nur für den Konzernabschlussprüfer, sondern auch für dessen verbundene Unternehmen angegeben sein dürften. Der enger gefasste Prüferbegriff nach dem BilMoG stimmt hingegen mit

⁸⁹ Vgl. Erwägungsgrund 7 der EU-VO.

⁹⁰ Vgl. z.B. Ratzinger-Sakel/Schönberger, *Accounting in Europe* 2015 S. 62 f. und 78 f., sowie Pott/Schröder/Weckelmann, *KoR* 2014 S. 502.

der hier relevanten Abgrenzung der EU-VO überein.⁹¹ Aufgrund der Empfehlung des IDW RS HFA 36 ist aber auch für die Zeit nach dem BilMoG nicht auszuschließen, dass teilweise nicht lediglich die Honorare für den Konzernabschlussprüfer angegeben sind.⁹² Eine Bereinigung der Honorare um den Anteil verbundener Unternehmen des Prüfers erfolgt insoweit nur, sofern der Anteil durch einen Davon-Vermerk kenntlich gemacht ist.⁹³ Eine Einschränkung der Auswertung ergibt sich dabei ganz generell auch dadurch, dass nicht sichergestellt ist, dass die Abgrenzung des Prüfers (und auch der Leistungsempfänger) über die vier Jahre, die für die Berechnung des Anteils der NPL an den durchschnittlichen APL jeweils relevant sind, durch die Unternehmen in gleicher Weise erfolgt. Außerdem ist zu berücksichtigen, dass dasselbe Fehler- oder dasselbe Vergleichsunternehmen für unterschiedliche Jahre jeweils mehrfach in der Fehler- bzw. Kontrollgruppe enthalten sein kann,⁹⁴ wobei aufgrund der im Folgenden vorzunehmenden zeitlichen Einschränkung der Stichprobe grds. nur Doppelungen zu beobachten sind.⁹⁵

Was die Abgrenzung der Leistungsempfänger im Konzern betrifft,⁹⁶ so ist im Fall eines Teilkonzernabschlusses die Abgrenzung nach § 314 Abs. 1 Nr. 9 HGB über den gesamten Zeitraum insofern enger als nach der EU-VO, als diese, wie oben erläutert, zumindest auch die Honorare für Leistungen an das übergeordnete Mutterunternehmen miteinschließt.⁹⁷ Für die Zeit nach dem BilMoG ist hingegen zu beachten, dass Honorare für Leistungen an Gemeinschaftsunternehmen in Deutschland quotal miteinbezogen werden sollen, nach Art. 4 Abs. 2 EU-VO hingegen nicht.⁹⁸ Was schließlich die Aufschlüsselung der Honorare in APL und NPL gem. EU-VO betrifft, so ist davon auszugehen, dass die der Kategorie der APL gem. § 314

⁹¹ Trotzdem wird im Folgenden auf eine zusätzliche getrennte Auswertung der Fälle vor und nach Inkrafttreten des BilMoG verzichtet, weil aufgrund des für die Berechnung des cap erforderlichen dreijährigen Vorlaufs für eine Auswertung nach Inkrafttreten des BilMoG nur Fälle ab dem Jahr 2012 hätten berücksichtigt werden können, sodass hier dann nur eine relativ kleine Stichprobe verblieben wäre.

⁹² Vgl. hierzu die Ausführungen in Teil 1, Abschn. 2.3.

⁹³ Fälle, bei denen die Honorare ausschließlich für einen länderübergreifenden Konzern angegeben sind, dem der Konzernabschlussprüfer angehört, werden somit ebenfalls nicht aus der folgenden Analyse ausgeschlossen. Fälle, bei denen explizit ausschließlich Honorardaten für das Netzwerk des Prüfers angegeben sind, werden hingegen nicht in die Auswertung einbezogen.

⁹⁴ Zur Begründung dieser Vorgehensweise vgl. Teil 1, Abschn. 2.3.

⁹⁵ Diese Doppelungen treten in der Kontrollgruppe im Folgenden deutlich häufiger auf als in der Fehlergruppe.

⁹⁶ Vgl. hierzu hinsichtlich § 314 Abs. 1 Nr. 9 HGB die Ausführungen in Teil 1, Abschn. 2.3.

⁹⁷ Vereinzelt gehen im Folgenden in die Auswertung für die Jahre vor dem Jahr des Fehlers auch Honorar- bzw. Prüferdaten aus Einzelabschlüssen ein. Hierbei handelt es sich aber durchweg um Fälle, bei denen das jeweilige Unternehmen im jeweiligen Jahr noch keinen Konzernabschluss erstellt hat, weil es noch nicht über Tochterunternehmen verfügte oder diese als unwesentlich eingestuft wurden.

⁹⁸ Letzteres gilt zumindest, sofern es sich nicht um APL im Rahmen der Prüfung der konsolidierten Abschlüsse handelt. Es gilt also insb. für alle NPL.

Abs. 1 Nr. 9 HGB zuzuordnenden Arten von Prüfungsleistungen nicht in jedem Fall mit den APL gem. Art. 4 Abs. 2 EU-VO übereinstimmen⁹⁹ und die der zweiten bis vierten Kategorie zuzuordnenden Leistungsarten nicht völlig deckungsgleich mit den durch den cap begrenzten NPL sind. Letzteres ist besonders offensichtlich für die in Art. 4 Abs. 2 EU-VO explizit ausgenommenen NPL, die nicht auf der black list stehen und nach EU-Recht oder nationalem Recht erforderlich sind. Diese werden nach § 314 Abs. 1 Nr. 9 HGB zumeist der zweiten Kategorie der anderen Bestätigungsleistungen (aBL) zuzuordnen sein.¹⁰⁰

Tab. 2.5: Überschreitung des cap im Jahr des Fehlers

	Fehler- gruppe	Kontroll- gruppe
Ausgangsstichprobe	151	151
abzüglich Fälle mit Bezug zu vor dem 01.01.2008 beginnenden Geschäftsjahren	65	65
Fälle mit Bezug zu am oder nach dem 01.01.2008 beginnenden Geschäftsjahren	86	86
abzüglich Fälle mit unvollständigen Honorarangaben	3	8
abzüglich Fälle mit Honorarangaben ausschließlich für das Prüfernnetzwerk	1	0
abzüglich Fälle ohne „Match“	10	6
Stichprobe für cap-Ermittlung	72	72
abzüglich Fälle ohne gleichen (Konzern-)Abschlussprüfer in den letzten 3 Jahren und im Jahr des Fehlers	17	21
abzüglich Fälle ohne NPL für letzte 3 aufeinanderfolgende Jahre	20	9
abzüglich Fälle mit NPL < cap	27	28
Anzahl der Fälle, in denen der cap überschritten wird (NPL > cap)	8	14
Anteil der Fälle, in denen der cap überschritten wird (NPL > cap)	11,1%	19,4%

Die näherungsweise Berechnung des cap kann im Folgenden nur für Geschäftsjahre vorgenommen werden, die am oder nach dem 01.01.2008 beginnen.¹⁰¹ Wie Tab. 2.5 entnommen werden kann, sind darüber hinaus aufgrund unvollständiger Daten weitere Ausschlüsse aus der Stichprobe vorzunehmen,¹⁰² sodass die für die Ermittlung des cap über den gesamten Zeitraum verbleibende

⁹⁹ Vgl. hierzu insb. IDW, a.a.O. (Fn. 88), S. 45-49, sowie Verlautbarung Nr. 4 (ü.F.) der APAS vom 20.12.2018 – Informationspflicht nach Art. 14 der Verordnung (EU) Nr. 537/2014 i.V.m. Fragen und Antworten zu Verlautbarung Nr. 4/1 der APAS vom 20.03.2019 – Hinweise zur Verlautbarung Nr. 4 (ü.F.), Art. 4 Abs.2, 13 Abs. 2 und Art. 14 AP-VO.

¹⁰⁰ Vgl. hierzu ausführlich IDW, a.a.O. (Fn. 88), S. 49 f. i.V.m. IDW RS HFA 36, Rdn. 13.

¹⁰¹ Dies ist darauf zurückzuführen, dass für die Berechnung der Obergrenze neben dem Honorar des jeweils betrachteten Jahres auch die Honorare der jeweils letzten drei Geschäftsjahre erforderlich sind und, wie in Teil 1, Abschn. 2.3, erläutert, die Pflicht zur Angabe der Honorare erstmals für Geschäftsjahre greift, die nach dem 31.12.2004 beginnen.

¹⁰² Aufgrund des durchgeführten 1-1-Match wird bei fehlenden Daten in der einen Gruppe auch jeweils die zugehörige Unternehmens-Jahr-Beobachtung aus der anderen Gruppe entfernt. Schließlich ist noch zu beachten, dass in jeder Gruppe auch jeweils noch zwei Fälle auszuschließen sind, bei denen zwar das Geschäftsjahr des jeweiligen

auswertbare Stichprobe der Fehlermeldungen und die zugehörige Kontrollstichprobe jeweils 72 Fälle umfassen.

Zu beachten ist, dass für die NPL nach Art. 4 Abs. 2 EU-VO überhaupt nur dann eine Honorarobergrenze besteht, wenn aus Sicht des Jahres des Fehlers das jeweilige Unternehmen in den letzten drei vorhergehenden Jahren durchgängig NPL von demselben Abschlussprüfer bezogen und denselben Prüfer auch im Jahr des Fehlers noch beibehalten hat. Wie Tab. 2.5 zu entnehmen ist, kommt es in der Fehlergruppe in 17 Fällen (23,6% der 72 Fälle) und in der Kontrollgruppe in 21 Fällen (29,2% von 72) zu einem Prüferwechsel im einschlägigen Zeitraum.¹⁰³ Bei den dann jeweils noch verbleibenden Fällen mit durchgängig gleichem Prüfer werden bei 20 Fällen der Fehlergruppe (27,8% von 72) und bei neun Fällen der Kontrollgruppe (12,5% von 72) in den drei Jahren vor dem Fehlerjahr nicht durchgängig NPL erbracht. Letztlich wird der cap somit nur in 11,1% der Fälle der Fehlergruppe überschritten. Nur in diesen acht von 72 Fällen hätte der Prüfer des fehlerhaften Konzernabschlusses bzw. -lageberichts also tendenziell den Umfang der NPL im Fehlerjahr aufgrund des Überschreitens der Honorargrenze einschränken müssen, wenn die EU-VO bereits gegolten hätte. Dabei wird von der in Teil 1, Abschn. 2.4, beschriebenen black list und deren Umsetzung in deutsches Recht noch ganz abstrahiert, die tendenziell ohnehin bereits zu einer Reduktion der NPL und damit möglicherweise zu einer weiteren Verringerung der Fälle mit überschrittener Honorargrenze geführt hätte. Der Anteil der Fälle, bei denen der cap überschritten

Unternehmens aus der Fehler- oder Kontrollgruppe erst am 01.01.2008 beginnt, der jeweilige Matchpartner jedoch ein vom Kalenderjahr abweichendes Geschäftsjahr mit Beginn vor dem 01.01.2008 aufweist.

¹⁰³ Zu beachten ist, dass ein Prüferwechsel im vorliegenden Zusammenhang nicht bereits dann angenommen wird, wenn eine andere Wirtschaftsprüfungsgesellschaft das Folgemandat übernimmt. Vielmehr wird für die Zwecke von Teil 3 dieses Beitrags ein Prüferwechsel enger abgegrenzt und diese Definition auch für diesen Abschnitt beibehalten; vgl. hierzu Teil 3, Abschn. 2.10.

worden wäre,¹⁰⁴ ist in der Kontrollgruppe mit 19,4% von 72 Fällen größer als in der Fehlergruppe.¹⁰⁵

Die deskriptive Auswertung liefert somit keine Hinweise dafür, dass der aus EU-Sicht kritische Level von 70% bei Fällen mit tendenziell niedriger Prüfungsqualität häufiger überschritten wird als bei Fällen mit tendenziell hoher Prüfungsqualität. Bei 88,9% der 72 untersuchten Fälle mit fehlerhaftem Konzernabschluss bzw. -lagebericht wäre die 70%-Honorargrenze gar nicht zum Tragen gekommen. Dies unterstützt die in der Literatur geäußerten Zweifel daran, dass die Einführung des cap in der gewählten Ausgestaltung eine besonders wirksame Maßnahme zur Steigerung der Prüfungsqualität ist.¹⁰⁶ Immerhin zeigt eine Betrachtung der 35 Fälle der Fehlergruppe, bei denen überhaupt eine Berechnung des cap hätte vorgenommen werden müssen, dass hier im Durchschnitt der errechnete Anteil der NPL an den durchschnittlichen APL der letzten drei Jahre bei 63,8% und damit nicht sehr weit unterhalb der 70%-Grenze liegt. Dies ist aber auf wenige Ausreißer zurückzuführen,¹⁰⁷ der Median liegt nämlich bei lediglich 44,6%

¹⁰⁴ Vgl. in diesem Zusammenhang auch eine Auswertung der Honorarangaben gem. § 314 Abs. 1 Nr. 9 HGB für 2013 bzw. 2013/2014 für 27 der 30 DAX-Unternehmen bei Ratzinger-Sakel/Schönberger, *Accounting in Europe 2015* S. 68 f. Dort wird für 33,3% der 27 betrachteten Unternehmen ein Überschreiten des cap festgestellt, wobei dieser Wert jedoch möglicherweise niedriger wäre, wenn die Fälle mit Prüferwechsel oder ohne durchgängige NPL über die letzten drei Jahre herausgerechnet würden. Bei Hohenfels/Quick, *Review of Managerial Science OnlineFirst* 25.10.2018, wird für 663 (25,8%) der 2.567 Unternehmens-Jahr-Beobachtungen des CDAX im Zeitraum 2006-2013 ein Überschreiten des cap festgestellt. Fälle mit Prüferwechsel oder ohne durchgängige NPL über die letzten drei Jahre scheinen dabei nicht herausgerechnet zu werden. Holm, *Arbeitspapier 2016*, verfügbar bei SSRN: <http://hbfm.link/4839>, Table 4 i.V.m. S. 23, ermittelt für 10,7% von 903 Unternehmens-Jahr-Beobachtungen für Deutschland im Zeitraum 2010-2013 ein Überschreiten des 70%-Anteils von NPL an APL, wobei hier jedoch offensichtlich zur Berechnung des Anteils die NPL eines Jahres lediglich durch die APL desselben Jahres dividiert werden. In gleicher Weise wird auch vorgegangen bei Köhler/Theis, *BFuP* 2018 S. 407 f., wo u.a. die Relevanz des cap für eine Stichprobe von 4.442 Unternehmens-Jahr-Beobachtungen für Deutschland im Zeitraum 2005-2015 abgeschätzt wird. Während außerhalb des DAX30 das durchschnittliche Verhältnis von NPL zu APL in allen Jahren unter 60% liegt, liegt dieses bei den untersuchten Unternehmen des DAX30 ab 2011 in jedem Jahr bei mehr als 70%.

¹⁰⁵ Einschränkend ist in diesem Zusammenhang allerdings anzumerken, dass sich unter den zehn Unternehmen der Fehlergruppe, die aufgrund ihres jeweils nicht auswertbaren Matchpartners aus der Analyse ausgeschlossen werden, zusätzlich fünf befinden, bei denen der cap überschritten worden wäre. Bei den sechs Unternehmen der Kontrollgruppe, die mangels Matchpartner nicht in die Analyse eingehen, wäre der cap hingegen in jedem Fall eingehalten worden.

¹⁰⁶ Vgl. in diesem Zusammenhang insb. auch die in die gleiche Richtung weisende Schlussfolgerung bei Hohenfels/Quick, *Review of Managerial Science OnlineFirst* 25.10.2018. Vgl. des Weiteren auch Lenz, *DB* 2016 S. 2558, der zu Recht darauf hinweist, dass die in Art. 4 Abs. 2 EU-VO vorgegebene Prüferabgrenzung, die im AReG nicht ausgeweitet wurde, den Abschlussprüfern erlaubt, die 70%-Honorargrenze „leerlaufen“ zu lassen, indem NPL auf verbundene Unternehmen oder Netzwerkunternehmen verlagert werden. Dabei wird auch auf eine insoweit strengere Umsetzung der EU-VO in Großbritannien hingewiesen. Vgl. ohne den zusätzlichen Verweis auf Großbritannien ebenso Schüppen, *VMEBF e.V., Stellungnahme zum AReG vom 14.02.2016*, S. 6, verfügbar unter: <http://hbfm.link/1026>, letzter Abruf 07.05.2019.

¹⁰⁷ So beträgt in einem Fall der errechnete Anteil nämlich sogar 213,5%. Dabei ist zu berücksichtigen, dass es unter den Fällen der Fehlergruppe einen weiteren Fall mit einem hohen Anteil von 198,3% gibt, bei dem in den aBL

und damit unterhalb des Medians der entsprechenden 42 Fälle der Kontrollgruppe, der bei 48,1% liegt. Der Durchschnitt für diese 42 Fälle ist hingegen knapp acht Prozentpunkte niedriger als in der Fehlergruppe und beträgt 55,9%. Hierbei ist aber zu beachten, dass bei dieser Berechnung etwas mehr als die Hälfte der betrachteten 72 Fälle der Fehlergruppe (und 41,7% der 72 Fälle der Kontrollgruppe) schon von vornherein ausgenommen sind, weil hier aufgrund eines Prüferwechsels bzw. eines nicht ununterbrochenen Bezugs von NPL im einschlägigen Zeitraum der cap gar nicht hätte berechnet werden müssen.¹⁰⁸

Tab. 2.6: Paarweiser Vergleich der Einhaltung des cap im Jahr des Fehlers^{a)}

		Kontrollgruppe		Σ
		> 70%	<= 70%	
Fehlergruppe	> 70%	2	6	8
	<= 70%	12	52	64
Σ		14	58	72

^{a)} Die Bezeichnung „<=70%“ schließt auch die Fälle ein, bei denen der cap gem. Art. 4 Abs. 2 EU-VO aufgrund eines Prüferwechsels oder einer Unterbrechung des Bezugs von NPL in den letzten drei Jahren gar nicht hätte berechnet werden müssen.

Betrachtet man abschließend unter expliziter Berücksichtigung des 1-1-Match¹⁰⁹ das Ergebnis der paarweisen Analyse in der Vier-Felder-Tafel von Tab. 2.6, so zeigt sich, dass sich lediglich 18 der 72 Paare im Hinblick auf die Einhaltung des cap unterscheiden. Die Wahrscheinlichkeit dafür, dass bei einem nicht übereinstimmenden Paar ein Unternehmen, bei dem der cap überschritten wird, ein Unternehmen mit Fehlermeldung ist, beträgt dabei lediglich 33,33% ($6 / (6 + 12) \times 100$) und ist damit halb so groß (odds ratio = $6 / 12 = 0,5$) wie die Wahrscheinlichkeit von 66,67%, dass es sich bei diesem Unternehmen um ein Unternehmen ohne Fehlermeldung handelt. Die Unterschiede

Versicherungsgebühren im Zusammenhang mit Comfort Letters enthalten sind. Diese wurden der Neuregelung des IDW RS HFA 36 n.F., Tz. 10, folgend, für die Auswertung herausgerechnet. Hätte man hierauf verzichtet, so wären der Anteil der NPL an den durchschnittlichen APL in diesem Fall auf 509,2% und der Durchschnitt für alle 35 Fälle auf 72,7% gestiegen.

¹⁰⁸ Streng genommen geht die bisherige Betrachtung im Übrigen von der Fiktion aus, dass die verbleibenden 35 (42) Fälle für die Fehlergruppe (Kontrollgruppe) tatsächlich durchgängig über die jeweils relevanten vier Jahre auch PIE waren. Ob diese Fiktion zutrifft, wird abschließend näherungsweise geprüft, indem für die jeweils relevanten Jahre überprüft wird, ob das Unternehmen auf den einschlägigen Listen der BaFin verzeichnet ist, die die dem Enforcement der DPR bzw. der BaFin unterliegenden Unternehmen enthalten; vgl. hierzu <http://hbfm.link/4832>, letzter Abruf: 30.05.2018. Die älteren Listen wurden auf Anfrage unmittelbar von der BaFin bezogen. Auf diese Weise wird jeweils ein Unternehmen der 35 Fehler- und der 42 Vergleichsunternehmen identifiziert, das nicht bereits die kompletten vier Jahre kapitalmarktorientiert war. Bei beiden Unternehmen wird der cap nicht überschritten.

¹⁰⁹ Vgl. hierzu Cram/Karan/Stuart, Contemporary Accounting Research 2009 insb. S. 479 ff., wo auf die Notwendigkeit hingewiesen wird, für eine korrekte statistische Analyse mit diskreter abhängiger Variable das Matching mittels einer bedingten logistischen Regression explizit zu berücksichtigen. Vgl. zur methodischen Vorgehensweise bei der folgenden Auswertung mittels univariater bedingter logistischer Regression auch Hosmer/Lemeshow/Sturdivant, Applied Logistic Regression, 3. Aufl. 2013, Kap. 7.

zwischen Fehler- und Kontrollgruppe hinsichtlich der Häufigkeit des Überschreitens des cap sind bei zweiseitigem Test zwar nicht signifikant,¹¹⁰ das Ergebnis der bedingten logistischen Regression liefert damit aber auch keinen Hinweis dafür, dass bei einem Überschreiten des cap eher mit einem Unternehmen mit Fehlermeldung und damit tendenziell niedriger Prüfungsqualität zu rechnen wäre.

2.7.2 Ist die 15%-Honorargrenze der EU-VO effektiv im Hinblick auf die Steigerung der Prüfungsqualität?

Das AReG hat die für alle Abschlussprüfer bereits zuvor in Deutschland gem. § 319 Abs. 3 Satz 1 Nr. 5 HGB geltende Honorargrenze zur Reduzierung der Gefahr von Umsatzabhängigkeit unverändert beibehalten. Danach ist ein Wirtschaftsprüfer immer dann von der Abschlussprüfung ausgeschlossen, wenn er in den letzten fünf Jahren und voraussichtlich auch im laufenden Jahr von der zu prüfenden KapGes. oder von Unternehmen, an denen diese mehr als 20% der Anteile hält, jeweils mehr als 30% der Gesamteinnahmen aus seiner beruflichen Tätigkeit bezieht. Dabei sind Einnahmen von Personen, mit denen der Abschlussprüfer seinen Beruf gemeinsam ausübt, in die Berechnung miteinzubeziehen.¹¹¹ Im Fall kapitalmarktorientierter Mandanten lag der zulässige Prozentsatz der Gesamteinnahmen gem. § 319a Abs. 1 Satz 1 Nr. 1 HGB a.F. bisher hingegen niedriger, nämlich bei 15%. Diese Regelung wurde durch das AReG zugunsten des unmittelbar auf PIEs anwendbaren Art. 4 Abs. 3 EU-VO aufgehoben.

Gem. Art. 4 Abs. 3 EU-VO muss der (Konzern-)Abschlussprüfer den Prüfungsausschuss der PIE in Kenntnis setzen, wenn er in den letzten drei aufeinanderfolgenden Geschäftsjahren jeweils mehr als 15% seiner insgesamt vereinnahmten Honorare von der PIE bezogen hat. Der Prüfungsausschuss diskutiert mit ihm die Gefahren für die Prüferunabhängigkeit und die zur Verminderung der Gefahren ergriffenen Schutzmaßnahmen. Er kann eine auftragsbegleitende Qualitätssicherungsprüfung durch einen anderen Abschlussprüfer vor Erteilung des

¹¹⁰ Der mittels einer univariaten bedingten logistischen Regression geschätzte Koeffizient, wenn als unabhängige Variable die dichotome Variable D (Nichteinhaltung des cap: D = 1; Einhaltung des cap: D = 0) und als abhängige Variable das Auftreten einer Fehlermeldung verwendet wird, beträgt -0,693. Der zugehörige p-Wert bei zweiseitigem Test beträgt 0,166. Der Koeffizient ist also nicht signifikant von Null verschieden. Eine Sensitivitätsanalyse, bei der alternativ das Matchingkriterium „Unternehmensgröße“ bzw. die logarithmierte Unternehmensgröße zusätzlich als unabhängige Variable berücksichtigt wird, führt hinsichtlich des Vorzeichens des Koeffizienten der dichotomen Variable und der Größenordnung des p-Werts zu keinem abweichenden Ergebnis. Fragt man hingegen, wie die Nichteinhaltung des cap vom Auftreten einer Fehlermeldung (Zugehörigkeit zur Fehlergruppe) und der nicht logarithmierten Unternehmensgröße abhängt, so ist der negative Koeffizient des Auftretens einer Fehlermeldung sogar schwach signifikant von Null verschieden (p = 0,085).

¹¹¹ Vgl. Schmidt/Nagel, in: Grottel u.a. (Hrsg.), Beck Bil-Komm., 11. Aufl. 2018, § 319 HGB Rdn. 70 i.V.m. Rdn. 31 f.

Bestätigungsvermerks veranlassen. Bei andauerndem Überschreiten der 15%-Honorargrenze soll der Prüfungsausschuss entscheiden, ob der bisherige (Konzern-)Abschlussprüfer die Prüfung für maximal zwei weitere Jahre erbringen darf. Anders als der vor dem 17.06.2016 in Deutschland geltende § 319a Abs. 1 Satz 1 Nr. 1 HGB a.F. verweist Art. 4 Abs. 3 EU-VO für die Berechnung des Prozentsatzes allein auf die Honorare, die der Prüfer von der PIE erhält, sodass nach dem Wortlaut der Vorschrift in die Berechnung weder Honorare von Tochter- noch Mutterunternehmen der PIE einzubeziehen wären. Hinsichtlich der Tochterunternehmen der PIE ergibt sich hier allerdings ein Widerspruch zu der einschlägigen in Erwägungsgrund 7 der EU-VO gegebenen Begründung. Auch aufseiten des Prüfers wird in Art. 4 Abs. 3 EU-VO, ebenso wie bei der Berechnung des cap gem. Art. 4 Abs. 2 EU-VO, lediglich auf die Gesamthonorare des Prüfers, nicht jedoch zusätzlich auch die Honorare seiner Tochter- oder Mutterunternehmen oder gar Netzwerkmitglieder abgehoben.¹¹² Art. 4 Abs. 4 EU-VO gilt auch i.V.m. Art. 4 Abs. 3 EU-VO, sodass die Mitgliedstaaten jeweils über die beschriebenen Anforderungen hinausgehende strengere Vorschriften erlassen können, wovon der deutsche Gesetzgeber jedoch genau wie beim cap im AReG abgesehen hat.

Die Regelung des Art. 4 Abs. 3 EU-VO zielt darauf ab, die (finanzielle) Abhängigkeit des Prüfers von einem Mandanten zu reduzieren und damit die Prüfungsqualität zu erhöhen.¹¹³ Im Folgenden soll deshalb unter Heranziehung der in Teil 1, Abschn. 2.3, beschriebenen Honorarangaben nach § 314 Abs. 1 Nr. 9 HGB zunächst für die Stichprobe mit Fehlerveröffentlichungen und damit Fällen mit tendenziell niedriger Prüfungsqualität überprüft werden, in wie vielen Fällen die bei der jeweiligen PIE erzielten Gesamthonorare des Konzernabschlussprüfers für das Jahr des Fehlers überhaupt so hoch gewesen wären, dass der von der EU offensichtlich als kritisch angesehene Umsatzanteil von 15% überschritten worden wäre. Zum Vergleich wird in einem zweiten Schritt dann für die Fälle der Kontrollstichprobe mit tendenziell hoher Prüfungsqualität die gleiche Auswertung durchgeführt, um zu überprüfen, ob der Anteil der Fälle, bei denen die 15%-Honorargrenze überschritten worden wäre, hier niedriger ist. Schließlich wird analog zur Vorgehensweise in Abschn. 2.7.1 und in Teil 1, Abschn. 2.4, in einer abschließenden Analyse der 1-1-Match explizit berücksichtigt.

¹¹² Vgl. hierzu insb. auch Verlautbarung Nr. 6 der APAS vom 10.04.2019 – Honorarabhängigkeit nach Art. 4 Abs. 3 der Verordnung (EU) Nr. 537/2014.

¹¹³ Vgl. die Begründung in Erwägungsgrund 7 der EU-VO.

Analog zu Abschn. 2.7.1 ist auch hier zu beachten, dass die Berechnung der 15%-Honorargrenze gem. EU-VO nur angenähert werden kann, weil die zur Verfügung stehenden Honorarangaben nicht exakt die Größe liefern, die nach Art. 4 Abs. 3 EU-VO im Zähler der Berechnungsformel erforderlich wäre. Hinsichtlich der Abgrenzung des Prüfers gelten insoweit die gleichen Ausführungen wie in Abschn. 2.7.1. Insb. für die Zeit vor Inkrafttreten des BilMoG besteht somit die Gefahr, dass, anders als von der EU-VO für die Berechnung der 15%-Honorargrenze vorgesehen, auch Honorare an die verbundenen Unternehmen des Prüfers in die Honorarangaben im Konzernanhang miteinbezogen sind.¹¹⁴ Deshalb erscheint eine zusätzliche getrennte Auswertung der Fälle vor und nach Inkrafttreten des BilMoG zweckmäßig. Auch hinsichtlich der Leistungsempfänger im Konzern scheint die Abgrenzung in Art. 4 Abs. 3 EU-VO tendenziell enger zu sein als nach § 314 Abs. 1 Nr. 9 HGB.¹¹⁵ Da die Honorarangaben aus dem Konzernanhang im Zähler der Berechnungsformel zu erfassen sind, könnten die beschriebenen Unterschiede somit dazu führen, dass die im Folgenden ermittelte Anzahl der Fälle, bei denen die Honorargrenze im Jahr des fehlerhaften Abschlusses jeweils überschritten wird, tendenziell zu hoch ausgewiesen wird.

Die im Nenner der Berechnungsformel anzusetzenden, insgesamt von allen Mandanten des jeweiligen Konzernabschlussprüfers erzielten Honorare werden dem Transparenzbericht des Prüfers entnommen. Wie in Teil 1, Abschn. 2.3, bereits erläutert, musste ein derartiger Bericht erstmals zum 31.03.2008 erstellt werden. Die in diesem ersten Bericht von den Prüfungsgesellschaften anzugebenden Umsätze beziehen sich wiederum auf das vorausgegangene Kalenderjahr 2007 bzw. ggf. das in 2007 oder spätestens bis zum 31.03.2008 endende vom Kalenderjahr abweichende Geschäftsjahr des jeweiligen Konzernabschlussprüfers. Unterstellt man vereinfachend für die im Zähler der Berechnungsformel angesetzten Honorare, dass die im Anhang des Konzernabschlusses der PIE ausgewiesenen Honorare des Konzernabschlussprüfers jeweils erst im Folgejahr bei der Prüfungsgesellschaft in die Umsatzerlöse eingehen, so können also erstmals Honorarangaben aus den Konzernabschlüssen des Jahres 2006 herangezogen werden, um unter Rückgriff auf die Daten aus den Transparenzberichten den entsprechenden Honoraranteil zu

¹¹⁴ Vgl. hierzu ausführlich Abschn. 2.7.1 i.V.m. Teil 1, Abschn. 2.3. Bei der Einbeziehung der Honorare und ggf. deren Bereinigung im Fall von Davon-Vermerken wird für diesen Abschnitt genauso vorgegangen wie in Abschn. 2.7.1.

¹¹⁵ Vgl. hierzu hinsichtlich § 314 Abs. 1 Nr. 9 HGB die Ausführungen in Teil 1, Abschn. 2.3.

berechnen.¹¹⁶ Die auswertbare Stichprobe der Fehlermeldungen und die zugehörige Kontrollstichprobe, die für die Ermittlung des Honoraranteils verbleiben, verringern sich aus diesem Grund sowie aufgrund weiterer fehlender Daten, wie Tab. 2.7 zu entnehmen ist, auf jeweils 113 Fälle.

Um festzustellen, ob der jeweilige Prüfungsausschuss bzw. der gesamte Aufsichts- oder Verwaltungsrat,¹¹⁷ wenn Art. 4 Abs. 3 EU-VO bereits in Kraft gewesen wäre, für das Fehlerjahr hätte erwägen müssen, vor Erteilung des jeweiligen Bestätigungsvermerks eine auftragsbegleitende Qualitätssicherungsprüfung durch einen anderen Abschlussprüfer zu fordern, hätten jeweils die Umsätze des Fehlerjahres und der zwei dem Fehlerjahr unmittelbar vorausgehenden Geschäftsjahre des jeweiligen Prüfers betrachtet werden müssen. Da die Gesamthonorare aus den Transparenzberichten der Prüfer frühestens für in 2007 endende Geschäftsjahre zur Verfügung stehen, hätten folglich nur Fehlerjahre, die in 2009 oder später enden, betrachtet werden können. Dies hätte dazu geführt, dass keine Auswertung von Fällen vor Inkrafttreten des BilMoG mehr möglich gewesen und die auswertbare Stichprobe und Kontrollstichprobe im Vergleich zu Abschn. 2.7.1 noch einmal deutlich kleiner geworden wäre. Deshalb wird im Folgenden lediglich unter Heranziehung der für das Jahr des Fehlers im Konzernanhang jeweils ausgewiesenen Honorare des Konzernabschlussprüfers überprüft, ob die in Art. 4 Abs. 3 EU-VO offensichtlich für die Unabhängigkeit des Prüfers als kritisch angesehene Grenze von 15% überschritten wird.

¹¹⁶ Vgl. zu dieser Vorgehensweise bei der Berechnung der Umsatzanteile im Zusammenhang mit der Regelung des § 319a Abs. 1 Satz 1 Nr. 1 HGB a.F. bereits Quick/Sattler, ZfB 2011 S. 77 f. i.V.m. Endnummer 21. Vgl. ebenso die Vorgehensweise in einem Beispielfall zu Art. 4 Abs. 3 EU-VO bei Lenz, DB 2016 S. 2559. Bei vom Kalenderjahr abweichenden Geschäftsjahren beim Mandanten und/oder Konzernabschlussprüfer wird jeweils der Transparenzbericht herangezogen, der die Umsatzdaten für das Geschäftsjahr des Prüfers enthält, in das das Datum des Bestätigungsvermerks für den Konzernabschluss im Jahr des Fehlers jeweils fällt. In den Fällen, in denen zwar Transparenzberichte hätten vorliegen müssen, diese aber nicht mehr beschafft werden konnten, oder in denen die Transparenzberichte keine Umsätze oder ausschließlich Konzernumsatzdaten des Prüfers enthalten, wurde auch noch nach dem einschlägigen Jahresabschluss des Prüfers im Bundesanzeiger gesucht.

¹¹⁷ Gem. Art. 39 Abs. 2 Unterabs. 2 der RL 2014/56/EU zur Änderung der RL 2006/43/EG über Abschlussprüfungen von Jahresabschlüssen und konsolidierten Abschlüssen können die Mitgliedstaaten erlauben oder vorschreiben, dass die Aufgaben eines Prüfungsausschusses auch vom Aufsichts- bzw. Verwaltungsrat insgesamt wahrgenommen werden können. Entsprechend sieht § 324 HGB für Deutschland vor, dass lediglich kapitalmarktorientierte Unternehmen, die keinen Aufsichts- oder Verwaltungsrat besitzen, der die Voraussetzungen des § 100 Abs. 5 AktG erfüllen muss, grds. verpflichtet sind, einen Prüfungsausschuss einzurichten.

Tab. 2.7: Überschreitung der 15%-Honorargrenze im Jahr des Fehlers

	Fehler- gruppe	Kontroll- gruppe
Ausgangsstichprobe	151	151
abzüglich Fälle für die Jahre 2004, 2005 und ggf. 2005/06	28	28
abzüglich Fälle, für die die Umsätze des Prüfers fehlen	6	4
abzüglich Fälle mit Honorarangaben ausschließlich für das Prüfernnetzwerk	1	0
abzüglich Fälle ohne „Match“	3	6
Stichprobe für die Ermittlung der 15%-Honorargrenze	113	113
<i>davon vor BilMoG^{a)}</i>	48	49
<i>davon nach BilMoG</i>	65	64
abzüglich Fälle, in denen die 15%-Honorargrenze nicht überschritten wird	102	109
<i>davon vor BilMoG</i>	42	49
<i>davon nach BilMoG</i>	60	60
Anzahl der Fälle, in denen die 15%-Honorargrenze überschritten wird	11	4
<i>davon vor BilMoG</i>	6	0
<i>davon nach BilMoG</i>	5	4
Anteil der Fälle, in denen die 15%-Honorargrenze überschritten wird	9,7%	3,5%
<i>Anteil vor BilMoG</i>	12,5%	0,0%
<i>Anteil nach BilMoG</i>	7,7%	6,3%

^{a)} Dass sich die Fallzahl vor und nach Inkrafttreten des BilMoG in der Fehler- und Kontrollgruppe jeweils um einen Fall unterscheidet, ist auf vom Kalenderjahr abweichende Geschäftsjahre zurückzuführen.

Tab. 2.7 zeigt, dass in lediglich 11 (9,7%) von 113 Fällen der Fehlergruppe mit tendenziell niedriger Prüfungsqualität die im Anhang des Konzernabschlusses des Fehlerjahres angegebenen Honorare des Prüfers so hoch sind, dass sie die aus EU-Sicht kritische Grenze von 15% der von allen Mandanten des jeweiligen Konzernabschlussprüfers erzielten Honorare überschreiten. In der Kontrollgruppe liegt der entsprechende Anteil noch deutlich niedriger, bei vier (3,5%) von 113 Fällen wird die kritische Grenze von 15% überschritten. Bei den insgesamt ausgewerteten 226 Fällen wird die kritische Honorargrenze von 15% bei Fällen mit tendenziell niedriger Prüfungsqualität also tatsächlich deutlich häufiger überschritten als bei Fällen mit tendenziell hoher Prüfungsqualität. Dies gilt allerdings nur bezogen auf den Gesamtzeitraum. Betrachtet man nur die Fälle nach Inkrafttreten des BilMoG, wo die Prüferabgrenzung bei den Honorarangaben nach § 314 Abs. 1 Nr. 9 HGB, wie oben erläutert, die Abgrenzung nach Art. 4 Abs. 3 EU-VO eventuell besser approximiert, so ist der Unterschied nur noch gering. Auch bezogen auf den Gesamtzeitraum ist der Anteil in der Fehlergruppe allerdings nicht hoch, 90,3% der Fälle liegen unterhalb der Grenze. Der Durchschnitt aller ermittelten Honoraranteile liegt für die 113 Fälle der Fehlergruppe bei

lediglich 4% und für die 113 Fälle der Kontrollgruppe bei 2,8%. Der Median (das dritte Quartil) beträgt sogar nur 0,2% (1,4%) in der Fehler- und 0,1% (0,5%) in der Kontrollgruppe.¹¹⁸

Betrachtet man unter expliziter Berücksichtigung des 1-1-Match das Ergebnis der paarweisen Analyse,¹¹⁹ so zeigt sich, dass sich im Gesamtzeitraum lediglich 13 der 113 Paare im Hinblick auf die Einhaltung der 15%-Honorargrenze unterscheiden. Die Wahrscheinlichkeit dafür, dass bei einem nicht übereinstimmenden Paar ein Unternehmen, bei dem die 15%-Honorargrenze überschritten wird, ein Unternehmen mit Fehlermeldung ist, beträgt dabei 76,92% ($10 / (10 + 3) \times 100$) und ist damit mehr als dreimal so groß (odds ratio = $10 / 3 = 3,3$) wie die Wahrscheinlichkeit von 23,08%, dass es sich bei diesem Unternehmen um ein Unternehmen ohne Fehlermeldung handelt. Die Unterschiede zwischen Fehler- und Kontrollgruppe hinsichtlich der Häufigkeit des Überschreitens der 15%-Honorargrenze sind bei zweiseitigem Test auf dem 10%-Niveau signifikant.¹²⁰ Begrenzt man die Betrachtung hingegen ausschließlich auf die Fälle nach dem BilMoG, so lassen sich keine signifikanten Unterschiede mehr feststellen.¹²¹ Für den Gesamtzeitraum stellt Abb. 2.1 das Ergebnis des 1-1-Match für alle 113 Paare anschaulich dar. Dazu sind auf der rechten Seite die Umsatzanteile für alle 113 Fälle der Fehlergruppe angetragen und durch eine Linie mit dem entsprechenden Umsatzanteil ihres jeweiligen Matchpartners aus der Kontrollgruppe auf der linken Seite verbunden.

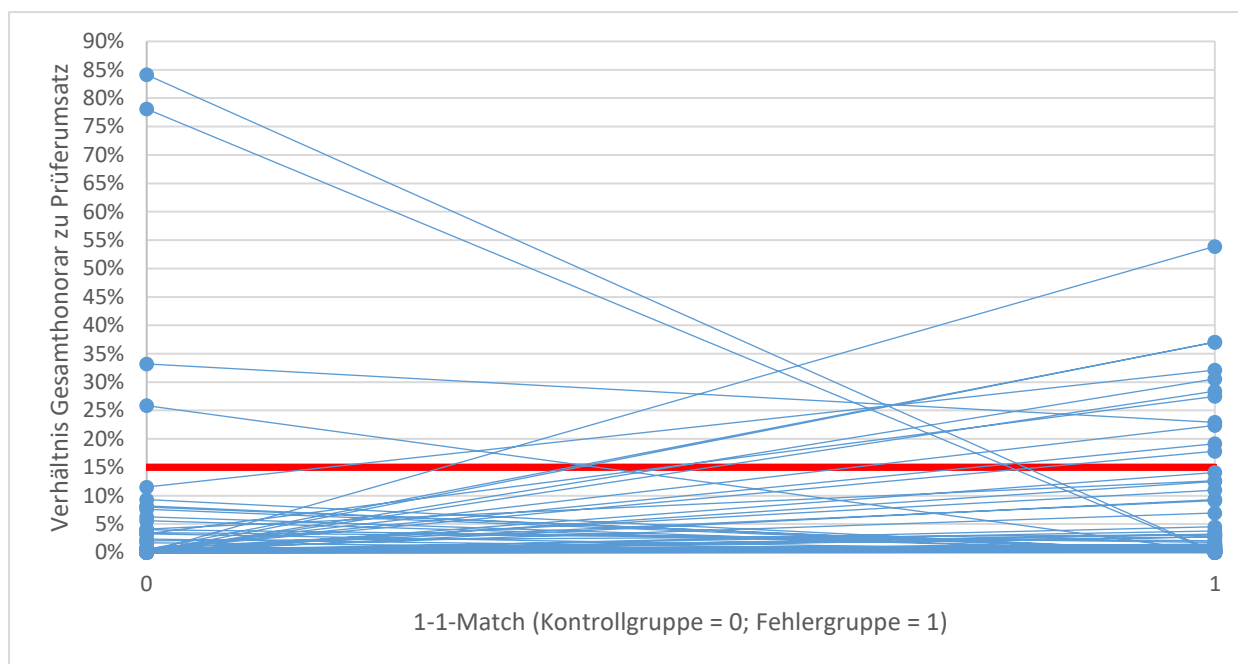
¹¹⁸ Maximal werden in der Fehlergruppe 53,9% und in der Kontrollgruppe 84,1% erreicht.

¹¹⁹ Vgl. hierzu bereits die Ausführungen und Literaturangaben in Abschn. 2.7.1, Fn. 109.

¹²⁰ Der mittels einer univariaten bedingten logistischen Regression geschätzte Koeffizient, wenn als unabhängige Variable die dichotome Variable D (Nichteinhaltung der 15%-Honorargrenze: $D = 1$; Einhaltung der 15%-Honorargrenze: $D = 0$) und als abhängige Variable das Auftreten einer Fehlermeldung verwendet wird, beträgt 1,204. Der zugehörige p-Wert bei zweiseitigem Test beträgt 0,067. Eine Sensitivitätsanalyse, bei der alternativ das Matchingkriterium „Unternehmensgröße“ bzw. die logarithmierte Unternehmensgröße zusätzlich als unabhängige Variable berücksichtigt wird, führt hinsichtlich Vorzeichen und Signifikanzniveau des Koeffizienten der dichotomen Variable zu keinem abweichenden Ergebnis. Fragt man hingegen, wie die Nichteinhaltung der 15%-Honorargrenze vom Auftreten einer Fehlermeldung (Zugehörigkeit zur Fehlergruppe) und der nicht logarithmierten Unternehmensgröße abhängt, so ist der positive Koeffizient des Auftretens einer Fehlermeldung sogar auf dem 5%-Niveau signifikant von Null verschieden ($p = 0,039$).

¹²¹ Es existieren in diesem Zeitraum lediglich sieben nicht übereinstimmende Paare, bei denen in vier Fällen beim Fehlerunternehmen und in drei Fällen beim Unternehmen aus der Kontrollgruppe die 15%-Honorargrenze überschritten wird.

Abb. 2.1: Paarweiser Vergleich der Einhaltung der 15%-Obergrenze im Jahr des Fehlers



Zu berücksichtigen ist auch, dass sich selbst bei den wenigen Fällen mit überschrittener Honorargrenze der hier errechnete, über 15% liegende Anteil aus den oben erläuterten Gründen lediglich auf die Honorare für das Jahr des Fehlers bezieht. Zur Klärung der Frage, ob der Prüfer, wenn Art. 4 Abs. 3 EU-VO bereits in Kraft gewesen wäre, vor Erteilung des Bestätigungsvermerks für das Fehlerjahr ggf. mit einer auftragsbegleitenden Qualitätssicherungsprüfung durch einen anderen Prüfer hätte rechnen müssen und eine erneute Mandatsübernahme im Folgejahr auch von einer Entscheidung des Prüfungsausschusses abhängig gewesen wäre, hätte die Überschreitung der Honorargrenze zusätzlich durchgängig auch für die drei vorhergehenden Konzernabschlüsse des Mandanten gelten müssen. Die bereits ohne diese zusätzliche Bedingung nicht sehr hohe Anzahl an Fällen mit einem Honoraranteil über 15% in der Fehlergruppe ist umso bemerkenswerter, wenn man die oben erläuterte Einschränkung berücksichtigt, dass aufgrund der verwendeten Honorarangaben gem. § 314 Abs. 1 Nr. 9 HGB der Honoraranteil im Vergleich zur Berechnungsvorschrift in Art. 4 Abs. 3 EU-VO eher überschätzt werden könnte.¹²²

¹²² Allerdings sind auch in diesem Abschnitt weitere Einschränkungen zu beachten. So ist zu berücksichtigen, dass dasselbe Fehler- oder dasselbe Vergleichsunternehmen für unterschiedliche Jahre jeweils mehrfach in der Fehler- bzw. Kontrollgruppe enthalten sein kann. Zur Begründung dieser Vorgehensweise vgl. Teil 1, Abschn. 2.3. Die beobachteten Umsatzanteile sind für diese Fälle möglicherweise nicht völlig unabhängig voneinander.

Auch hinsichtlich der 15%-Honorargrenze können also Zweifel daran nicht ausgeräumt werden, dass die EU-Regelung besonders effektiv ist, um breitflächig zu einer substanziellen Verbesserung der Prüfungsqualität zu gelangen.¹²³ Das, wenn auch nur über den Gesamtzeitraum, deutlich bessere Ergebnis in der Kontrollgruppe könnte immerhin als Hinweis für die grds. richtige Stoßrichtung der Regelung aufgefasst werden, die mit gleichem Prozentsatz, aber u.a. längerem Zeithorizont und anderer Abgrenzung von Prüfer und Leistungsempfänger für kapitalmarktorientierte Unternehmen auch vor Inkrafttreten der EU-VO durch § 319a Abs. 1 Satz 1 Nr. 1 HGB a.F. in Deutschland bereits umgesetzt war. Auch aus theoretischer Sicht kann argumentiert werden, dass es für die Frage der Höhe der finanziellen Abhängigkeit eines Prüfers aus einem bestimmten Mandat nicht allein auf das Ergebnis aus diesem Mandat, sondern auf den Anteil dieses Ergebnisses am Gesamtergebnis aus allen Mandaten des jeweiligen Prüfers ankommt.¹²⁴ Für die Wirksamkeit der Regelung ist aber offensichtlich deren konkrete Ausgestaltung entscheidend. So erscheint die vorgegebene Grenze angesichts der vorgelegten Bestandsaufnahme als hoch.¹²⁵ Aus ökonomischer Sicht erscheint auch nicht plausibel, dass eine Regelung geeignet sein soll, die auf Prüferseite nur auf den gesetzlichen (Konzern-)Abschlussprüfer abhebt und damit eine Umgehung der Regelung leicht ermöglicht, indem Leistungen für den Mandanten nicht durch den Prüfer selbst, sondern z.B. durch verbundene Unternehmen des Prüfers erbracht werden.¹²⁶

¹²³ Vgl. zu einer ähnlichen Schlussfolgerung im Hinblick auf die Effektivität der einschlägigen Regelung im HGB vor dem AReG bereits die empirische Studie für die Jahre 2005 bis 2011 von Lopatta/Kaspereit/Canitz/Maas, BFuP 2015 S. 582 f. i.V.m. S. 572. Dort wird für deutsche Unternehmen des DAX, MDAX, SDAX und TecDAX ein signifikant positiver Zusammenhang zwischen dem Umsatzanteil des Prüfers und positiven diskretionären working capital accruals abgeleitet, wobei der durchschnittliche Umsatzanteil nur ca. 0,7% beträgt. Keinen signifikanten Zusammenhang zwischen Umsatzanteil und diskretionären Periodenabgrenzungen als Surrogatmaß für Prüfungsqualität finden hingegen z.B. Quick/Sattler, ZfB 2011 S. 84 i.V.m. S. 67 f., deren Stichprobe bedeutende deutsche Unternehmen des Prime Standard von 2005 bis 2007 umfasst, sowie Krauß/Zülch, The Journal of Applied Business Research 2013 S. 321 i.V.m. S. 313, die den Zeitraum 2004 bis 2011 für deutsche Unternehmen des DAX, MDAX, SDAX und TecDAX betrachten.

¹²⁴ Vgl. hierzu ausführlich die modelltheoretischen Überlegungen bei DeAngelo, Journal of Accounting and Economics 1981 S. 187-192. Zu beachten ist allerdings, dass es aus dieser Sicht nicht auf den Anteil am Prüferumsatz, sondern auf den Anteil an den sog. Quasirenten ankommt, die jedoch kaum justiziabel gemessen werden können. Vgl. zu Letzterem Wagenhofer/Ewert, Externe Unternehmensrechnung, 3. Aufl. 2015, S. 525.

¹²⁵ Vgl. hierzu z.B. die strengere Umsetzung von Art. 4 Abs. 3 EU-VO in Großbritannien, wo eine Abschlussprüfungstätigkeit bei einer PIE bereits bei einem erwarteten Honoraranteil von regelmäßig mehr als 10% ausgeschlossen sein soll. Darüber hinaus soll bei einem entsprechenden Honoraranteil größer 5% und kleiner gleich 10% bereits mit dem Ethics Partner der Prüfungsgesellschaft sowie dem Prüfungsausschuss des Mandanten über die Notwendigkeit potenzieller Schutzmaßnahmen befunden werden; vgl. Financial Reporting Council (FRC), Revised Ethical Standards 2016, S. 71 ff., Rdn. 4.42, 4.43 und 4.47.

¹²⁶ Vgl. hierzu z.B. bereits die Überlegungen bei Lenz, DB 2016 S. 2557 ff.

2.8 Zusammenfassung zu Teil 2 und Ausblick

Die Verordnung (EU) Nr. 537/2014 über spezifische Anforderungen an die Abschlussprüfung bei PIEs vom April 2014 sieht u.a. eine Deckelung der Honorare vor, die der jeweilige (Konzern-)Abschlussprüfer maximal aus NPL erzielen darf (cap). Als weitere Honorargrenze beinhaltet die EU-VO auch eine Beschränkung des maximal zulässigen Anteils der Honorare aus Prüfung und NPL aus einem bestimmten PIE-Mandat am gesamten Umsatz des jeweiligen (Konzern-)Abschlussprüfers aus allen Mandaten. Beide Vorkehrungen sollen der Verbesserung der Prüfungsqualität dienen.

In Teil 2 des Beitrags werden anhand einer deskriptiven Analyse Hinweise zur Beantwortung der Frage geliefert, ob diese Vorgaben der EU-VO sowie die damit zusammenhängende Ausübung bzw. Nichtausübung in der EU-VO vorgesehener Mitgliedstaatenwahlrechte im Rahmen des AReG tatsächlich geeignet sind, zur Verbesserung der Abschlussprüfungsqualität beizutragen. Hierzu wird für eine Stichprobe von Fällen mit tendenziell niedriger Prüfungsqualität untersucht, inwieweit die durch die EU-VO vorgesehenen Beschränkungen hinsichtlich des Umfangs von Prüfungs- und Nichtprüfungsleistungen bei diesen Fällen überhaupt zum Tragen gekommen wären, wenn sie im Untersuchungszeitraum bereits zu beachten gewesen wären. Als Surrogatmaß für eine niedrige Prüfungsqualität werden dabei zwischen dem 01.07.2005 und dem 31.12.2017 im Bundesanzeiger veröffentlichte Meldungen von Rechnungslegungsfehlern verwendet, die von der DPR bzw. der BaFin im Rahmen eines sog. Enforcementverfahrens festgestellt wurden. Die Bestandsaufnahme wird jeweils mit dem Befund einer Kontrollstichprobe mit tendenziell hoher Prüfungsqualität verglichen.

Was die Vorschrift zum cap betrifft, die die Höhe der maximal zulässigen Honorare aus NPL an die Höhe der erzielten Honorare aus APL bindet, so liefert die Auswertung keine Hinweise dafür, dass der aus EU-Sicht kritische Level von 70% bei Fällen mit tendenziell niedriger Prüfungsqualität häufiger überschritten wird als bei Fällen mit hoher Prüfungsqualität. Im Übrigen wäre bei 88,9% der in diesem Zusammenhang untersuchten 72 Fälle mit fehlerhaftem Konzernabschluss bzw. -lagebericht die 70%-Honorargrenze gar nicht zum Tragen gekommen. Dies unterstützt die in der Literatur geäußerten Zweifel daran,¹²⁷ dass die Einführung des cap in

¹²⁷ Vgl. z.B. Ratzinger-Sakel/Schönberger, *Accounting in Europe* 2015 S. 62 f. und 78 f., sowie Pott/Schröder/Weckelmann, *KoR* 2014 S. 502.

der gewählten Ausgestaltung eine besonders wirksame Maßnahme zur Steigerung der Prüfungsqualität ist.¹²⁸

Auch hinsichtlich der in Art. 4 Abs. 3 EU-VO vorgesehenen 15%-Honorargrenze können Zweifel daran nicht ausgeräumt werden, dass diese EU-Regelung besonders effektiv ist, um breitflächig zu einer substantiellen Verbesserung der Prüfungsqualität zu gelangen. Das, wenn auch nur über den Gesamtzeitraum, deutlich bessere Ergebnis in der Kontrollgruppe könnte immerhin als Hinweis für die grds. richtige Stoßrichtung der Regelung gedeutet werden. So ergibt die paarweise Analyse von hinsichtlich des Überschreitens der Honorargrenze nicht übereinstimmenden Paaren aus Fehler- und Vergleichsunternehmen, dass die Wahrscheinlichkeit dafür, dass ein Unternehmen, bei dem die 15%-Honorargrenze für das Fehlerjahr überschritten wird, ein Unternehmen mit Fehlermeldung ist, mehr als dreimal so groß ist wie die Wahrscheinlichkeit, dass es sich bei diesem Unternehmen um ein Unternehmen ohne Fehlermeldung handelt. Für die Wirksamkeit der Regelung ist aber offensichtlich deren konkrete Ausgestaltung entscheidend. So erscheint die vorgegebene Grenze von 15% Umsatzanteil angesichts der vorgelegten Bestandsaufnahme als hoch. Aus ökonomischer Sicht erscheint schließlich auch nicht plausibel, dass eine Regelung geeignet sein soll, die Unabhängigkeit des Prüfers zu erhöhen, die auf Prüferseite nur auf den gesetzlichen (Konzern-)Abschlussprüfer abhebt und damit eine Umgehung der Regelung leicht ermöglicht, indem Leistungen für den Mandanten nicht durch den Prüfer selbst, sondern z.B. durch verbundene Unternehmen des Prüfers erbracht werden.¹²⁹ Im demnächst folgenden Teil 3 dieser empirischen Bestandsaufnahme wird vor dem Hintergrund der durch die EU-VO zusätzlich auch eingeführten externen Prüferrotation schließlich noch betrachtet werden, inwieweit in Fehler- und Kontrollgruppe eine freiwillige Prüferrotation beobachtet werden kann.

2.9 Einleitung zu Teil 3

Die Verordnung (EU) Nr. 537/2014 (EU-VO) über spezifische Anforderungen an die Abschlussprüfung bei Unternehmen von öffentlichem Interesse (PIEs) vom April 2014 hat mit dem Ziel der Verbesserung der Qualität der Abschlussprüfung bei PIEs u.a. die Erbringung von Nichtprüfungsleistungen (NPL) durch den Abschlussprüfer sowohl hinsichtlich der Kategorien

¹²⁸ Naturgemäß ist bei dieser und der folgenden Interpretation der Ergebnisse zu beachten, dass auch die vorliegende Studie wie jede empirische Untersuchung spezifischen Einschränkungen unterliegt. Diese werden in der Schlussbetrachtung des Gesamtbeitrags in Teil 3, Abschn. 2.11, zusammengefasst.

¹²⁹ Vgl. hierzu z.B. bereits die Überlegungen bei Lenz, DB 2016 S. 2557 ff. Analog wird dort auch im Hinblick auf die Regelung des Art. 4 Abs. 2 EU-VO argumentiert.

zulässiger NPL als auch im Hinblick auf den erlaubten Umfang (cap, zulässiger Umsatzanteil von Honoraren aus Prüfungs- und Nichtprüfungsleistungen) weiter eingeschränkt. Auch mit der erstmaligen verbindlichen Einführung einer externen Prüferrotation soll die Prüfungsqualität verbessert werden.¹³⁰

In der Literatur finden sich nun jedoch durchaus Zweifel, ob die konkreten Vorgaben der EU-VO tatsächlich geeignet sind, zur Verbesserung der Abschlussprüfungsqualität beizutragen.¹³¹ Vor diesem Hintergrund erfolgte in Teil 1 und 2 dieser Studie für eine Stichprobe von Fällen mit tendenziell niedriger Prüfungsqualität eine deskriptive Auswertung dahingehend, inwieweit die durch die EU-VO vorgesehenen Beschränkungen hinsichtlich Art und/oder Umfang von NPL bei diesen Fällen überhaupt zum Tragen gekommen wären, wenn sie im Untersuchungszeitraum bereits zu beachten gewesen wären. Die Ausübung bzw. Nichtausübung in der EU-VO vorgesehener Mitgliedstaatenwahlrechte im Rahmen des am 17.06.2016 in Kraft getretenen Abschlussprüfungsreformgesetzes (AReG) wurde dabei mit berücksichtigt. Die Bestandsaufnahme wurde jeweils mit dem Befund einer Kontrollstichprobe verglichen. Im vorliegenden Teil 3 der Studie wird schließlich noch betrachtet, inwieweit bei den Fällen mit tendenziell niedriger Prüfungsqualität eine freiwillige Prüferrotation zu beobachten war, wobei auch hier ein Vergleich mit einer Kontrollstichprobe erfolgt.¹³²

Als Surrogatmaß für eine niedrige Prüfungsqualität werden in dieser Untersuchung zwischen dem 01.07.2005 und dem 31.12.2017 im Bundesanzeiger veröffentlichte Meldungen von Rechnungslegungsfehlern verwendet, die von der Deutschen Prüfstelle für Rechnungslegung (DPR) bzw. der BaFin im Rahmen eines sog. Enforcementverfahrens festgestellt wurden. Teil 1 dieses Beitrags (Abschn. 2.1-2.5) hat die Grenzen des gewählten Surrogatmaßes für Prüfungsqualität aufgezeigt und eine Bereinigung und genaue Abgrenzung der Stichprobe der Fehlermeldungen (Fehlergruppe) vorgenommen. Der Fehlergruppe wurde mittels eines sog. 1-1-Match eine Kontrollstichprobe (Kontrollgruppe) zugeordnet.¹³³ Außerdem wurden weitere für die Untersuchung erhobene Daten erläutert. Teil 1 schließt mit einer deskriptiven Analyse der in

¹³⁰ Vgl. Erwägungsgründe 7, 8 und 21 der EU-VO.

¹³¹ Vgl. im Hinblick auf black list und cap z.B. Ratzinger-Sakel/Schönberger, *Accounting in Europe* 2015 S. 62 f.; im Hinblick auf die externe Rotation vgl. z.B. Pott/Schröder/Weckelmann, *KoR* 2014 S. 504.

¹³² Zur Abgrenzung der Kontrollstichprobe vgl. die Ausführungen in Teil 1, Abschn. 2.3.

¹³³ Ein 1-1-Match findet sich in mehreren bisherigen Studien zum deutschen Enforcement; vgl. z.B. Strohmenger, *Journal of International Financial Management & Accounting* 2014 S. 281, sowie Ebner/Hottmann/Zülch, *Corporate Ownership & Control* 2017 S. 128.

Fehler- und Kontrollgruppe auftretenden Kategorien von Prüfungs- und Nichtprüfungsleistungen. Hieraus kann unter dem Gesichtspunkt der Verbesserung der Prüfungsqualität kein Argument für die Verschärfung des Verbots von NPL durch die EU-VO abgeleitet werden. Teil 2 dieses Beitrags (Abschn. 2.6-2.8) widmete sich den von der EU-VO in quantitativer Hinsicht vorgesehenen Einschränkungen bei der Erbringung von Prüfungs- bzw. Nichtprüfungsleistungen. Dabei wurde zunächst die Deckelung der Honorare, die der jeweilige (Konzern-)Abschlussprüfer maximal aus NPL erzielen darf (sog. cap), betrachtet. Die Analyseergebnisse unterstützen Zweifel daran, dass die Einführung des cap in der gewählten Ausgestaltung eine besonders wirksame Maßnahme zur Steigerung der Prüfungsqualität ist. Abschließend widmete sich Teil 2 der Begrenzung des maximal zulässigen Anteils der Honorare aus Prüfung und NPL aus einem bestimmten PIE-Mandat am gesamten Umsatz des jeweiligen (Konzern-)Abschlussprüfers aus allen Mandaten. Zwar könnte das Ergebnis der zugehörigen Auswertung als Hinweis für die grds. richtige Stoßrichtung der Regelung interpretiert werden. Zweifel an deren Effektivität verbleiben jedoch. Teil 3 dieser Studie befasst sich nun mit der für die Abschlussprüfer von PIEs durch die EU-VO eingeführten Pflicht zur externen Rotation. Abschn. 2.10 führt in das Thema ein und analysiert die freiwilligen Prüferwechsel in der Fehler- und Kontrollgruppe. Abschn. 2.11 fasst die Ergebnisse von Teil 3 sowie die wesentlichen Befunde des Gesamtbeitrags zusammen und geht insb. ausführlich auf Einschränkungen ein, denen die empirische Bestandsaufnahme unterliegt.

2.10 Untersuchungsergebnisse – Externe Rotation

Vor Inkrafttreten der EU-VO bestand in Deutschland bei der Prüfung kapitalmarktorientierter Mandanten gem. § 319a Abs. 1 Satz 1 Nr. 4, Satz 4 und 5 HGB a.F. i.V.m. § 319a Abs. 2 HGB für den verantwortlichen Prüfungspartner nur die Pflicht zur internen Rotation. Dies bedeutet, dass dieser nach spätestens sieben Jahren die Verantwortung für die Prüfung an einen anderen Prüfungspartner, der jedoch bei derselben Wirtschaftsprüfungsgesellschaft beschäftigt sein konnte, abzugeben hatte und auch nicht mehr Mitglied des Prüfungsteams sein konnte.¹³⁴ Durch Art. 17 Abs. 1 EU-VO wurde für die Prüfung von PIEs die Pflicht zur externen Rotation eingeführt, die

¹³⁴ Vgl. zur Auslegung dieser Vorschrift z.B. Schmidt/Nagel in: Grottel u.a. (Hrsg.), Beck Bil-Komm., 10. Aufl. 2016, § 319a HGB Rdn. 31-39, sowie Ebke, in: Schmidt/Ebke (Hrsg.), Münchener Komm. zum HGB, Bd. 4, 3. Aufl. 2013, § 319a HGB Rdn. 24-32. Die Pflicht zur internen Rotation nach spätestens sieben Jahren ergibt sich zukünftig unmittelbar aus Art. 17 Abs. 7 EU-VO. Von einer möglichen Verkürzung dieser Frist hat der deutsche Gesetzgeber im AREG abgesehen. Zu den verpflichtenden Neuerungen der EU-VO im Vergleich zur aufgehobenen HGB-Regelung zählt die Verlängerung der Abkühlungsphase nach der internen Rotation von zwei auf drei Jahre sowie die Einführung eines graduellen Rotationssystems für das beteiligte Führungspersonal; vgl. Art. 17 Abs. 7 EU-VO.

grds. einen Wechsel des Abschlussprüfers und damit ggf. insb. auch der Prüfungsgesellschaft nach 10 Jahren vorsieht. Nach Ablauf dieser Höchstlaufzeit ist vor einer erneuten Mandatsübernahme bei derselben PIE gem. Art. 17 Abs. 3 EU-VO eine sog. Cooling-off-Periode von 4 Jahren einzuhalten, die auch für alle Netzwerkmitglieder des bisherigen Abschlussprüfers in der EU gilt. In Art. 17 Abs. 2 Buchst. b und Abs. 4 EU-VO wird den Mitgliedstaaten das Wahlrecht eingeräumt, die Höchstlaufzeit entweder zu verkürzen oder zu verlängern. Gem. Art. 17 Abs. 4 EU-VO kann sie bei Durchführung eines Auswahl- und Vorschlagsverfahrens nach Art. 16 Abs. 2 bis 5 EU-VO auf 20 Jahre angehoben werden. Werden Gemeinschaftsprüfungen, sog. joint audits, durchgeführt, ist eine Verlängerung auf 24 Jahre möglich.¹³⁵

Als Nachteil einer Pflicht zur externen Rotation wird aus theoretischer Sicht u.a. regelmäßig angeführt, dass in den ersten Jahren einer Prüfer-Mandanten-Beziehung die Urteilsfähigkeit des Prüfers aufgrund (noch) fehlender mandantenspezifischer Kenntnisse niedriger wäre und somit die Prüfungsqualität aufgrund der mit einem externen Prüferwechsel verbundenen Informationsverluste leiden könnte.¹³⁶ Um diese mit häufigen Prüferwechseln verbundene Gefahr möglichst gering zu halten, hat sich der deutsche Gesetzgeber mit § 318 Abs. 1a HGB für die volle Ausschöpfung des Mitgliedstaatenwahlrechts im Hinblick auf die Höchstlaufzeit eines Mandats entschieden.¹³⁷ Aufgrund ihrer besonderen Bedeutung¹³⁷ für den Finanzmarkt gilt dies allerdings nicht für Kreditinstitute und Versicherungsunternehmen.¹³⁸

Umgekehrt wird als Begründung für die Einführung einer externen Rotation u.a. regelmäßig angeführt, dass mit längerer Mandatslaufzeit die nach der Einarbeitung und des Erwerbs mandantenspezifischen Wissens gestiegene Urteilsfähigkeit durch zunehmende „Betriebsblindheit“ des Abschlussprüfers auch wieder sinken könnte.¹³⁹ Vor allem aber wird auf

¹³⁵ Gem. Art. 17 Abs. 6 EU-VO kann die Höchstlaufzeit auf Antrag der PIE von der zuständigen Behörde im Ausnahmefall um bis zu zwei weitere Jahre verlängert werden. Zuständige Behörde in Deutschland ist die Abschlussprüferaufsichtsstelle (APAS); vgl. IDW, Positionspapier zu Inhalten und Zweifelsfragen der EU-Verordnung und der Abschlussprüferrichtlinie, 4. Aufl. (Stand: 23.05.2018), S. 26.

¹³⁶ Vgl. z.B. Luik, BB 1976 S. 238 f., Herzig/Watrin, zfbf 1995 S. 795, sowie Quick/Wiemann, ZCG 2013 S. 79.

¹³⁷ Vgl. RegE AREG vom 16.12.2015, Begr. B. Besonderer Teil, zu Art. 1, zu Nr. 2 (§ 318 HGB), S. 45 f. Damit die Höchstlaufzeit greift, muss das Auswahl- und Vorschlagsverfahren mit Wirkung für das elfte Geschäftsjahr stattfinden, auf das sich die ununterbrochene Tätigkeit als Abschlussprüfer erstreckt, bzw. müssen die joint audits ab diesem elften Geschäftsjahr durchgängig erfolgen; vgl. insb. auch unmittelbar Art. 17 Abs. 4 Buchst. a und b EU-VO.

¹³⁸ Vgl. §§ 340k Abs. 1 Satz 1, 341k Abs. 1 Satz 2 HGB i.V.m. RegE AREG vom 16.12.2015, Begr. B. Besonderer Teil, zu Art. 1, zu Nr. 2 (§ 318 HGB), S. 45 f.

¹³⁹ Vgl. z.B. Wagenhofer/Ewert, Externe Unternehmensrechnung, 3. Aufl. 2015, S. 531, Fn. 33, sowie kritisch insoweit z.B. Luik, BB 1976 S. 238, und Quick/Wiemann, ZCG 2013 S. 79. Vgl. zu einer Zusammenstellung in der Literatur

Probleme hinsichtlich der Unabhängigkeit des Prüfers bei unbeschränkten Mandatslaufzeiten hingewiesen.¹⁴⁰ So sieht die EU die Unabhängigkeit insb. aufgrund der mit zunehmender Mandatslaufzeit anwachsenden Vertrautheit zwischen Prüfer und Mandant gefährdet (sog. familiarity threat).¹⁴¹

Vor diesem Hintergrund soll im Folgenden überprüft werden, ob hinsichtlich der im Jahr des Fehlers bereits abgelaufenen Mandatslaufzeit des jeweils im Amt befindlichen Konzernabschlussprüfers systematische Unterschiede zwischen den Fällen der Fehler- und der Kontrollgruppe bestehen. Nach den obigen Überlegungen wäre vor allem bei zu beobachtenden Mandatsdauern, die erst wenige Jahre oder bereits sehr viele Jahre umfassen, mit einer tendenziell niedrigeren Prüfungsqualität zu rechnen. Für kurze Mandatsdauern enthält Tab. 2.8 die Auswertung für die Fehler- und die Kontrollgruppe. Dabei wird zunächst untersucht, ob in der Fehlergruppe eine größere Anzahl von im Hinblick auf den jeweils betroffenen Konzern noch nicht erfahrenen Prüfern den fehlerhaften Konzernabschluss bzw. -lagebericht geprüft hat, als dies in der Kontrollgruppe für den jeweiligen Abschluss des Fehlerjahres der Fall ist. Die Analyse der Fehlergruppe ergibt gem. Tab. 2.8, dass in 12,6% aller 151 Fälle der für die Prüfung des fehlerhaften Konzernabschlusses bzw. -lageberichts zuständige Prüfer erstmals mit der Konzernabschlussprüfung bei dem jeweiligen Mandanten beauftragt war.¹⁴² Der entsprechende

diskutierter Vor- und Nachteile einer externen Pflichtrotation z.B. auch Marten/Quick/Ruhnke, *Wirtschaftsprüfung*, 5. Aufl. 2015, S. 192 ff.

¹⁴⁰ In diesem Zusammenhang wird in der Literatur auch immer wieder angeführt, dass das finanzielle Eigeninteresse an der Beibehaltung des Mandats und damit möglicherweise die Bereitschaft des Prüfers, nicht ordnungsgemäß Bericht zu erstatten, umso größer wäre, je länger der Prüfer das Mandat noch behalten könnte; vgl. hierzu z.B. Quick/Wiemann, *ZCG* 2013 S. 79. Vgl. demgegenüber aber z.B. Wagenhofer/Ewert, a.a.O. (Fn. 139), S. 529 f., wo darauf hingewiesen wird, dass eine solche, auf der Quasirentenmentation von DeAngelo, *Journal of Accounting and Economics* 1981 S. 115-123, aufbauende Folgerung hinsichtlich abnehmender finanzieller Abhängigkeit des Prüfers bei begrenzter Mandatslaufzeit dann nicht zwingend überzeugt, wenn man unterstellt, dass die Begrenzung der Mandatslaufzeit durch Einführung einer externen Rotation nicht nur bei einem bestimmten Mandat des Prüfers, sondern auch bei seinen anderen Mandaten zum Tragen kommt.

¹⁴¹ Vgl. Erwägungsgrund 21 der EU-VO. Zum Begriff des familiarity threat vgl. insb. IESBA *International Code of Ethics for Professional Accountants* 2018, 120.6 A3 d).

¹⁴² Zu beachten ist, dass ein Prüferwechsel für die Zwecke dieses Abschnitts nicht bereits dann angenommen wird, wenn eine Wirtschaftsprüfungsgesellschaft mit abweichendem Namen das Folgemandat übernimmt. Vielmehr wird ein Prüferwechsel enger abgegrenzt. Trotz Wechsels der Gesellschaft bzw. der Firma soll von keinem Prüferwechsel ausgegangen werden, wenn lediglich eine Umfirmierung vorliegt, vorhergehende und neue Prüfungsgesellschaft sich zusammengeschlossen haben (vgl. zu Letzterem insb. auch bereits Art. 17 Abs. 8 Unterabs. 2 EU-VO) oder das Mandat lediglich innerhalb des Konzerns oder auch nur innerhalb des Netzwerks der Prüfungsgesellschaft auf eine andere Gesellschaft übergeht. Überprüft wurde dies anhand der Transparenzberichte, es sei denn, eine Netzwerkverbindung war unmittelbar anhand der Namen der Gesellschaften erkennbar. Insb. wo Transparenzberichte (noch) nicht vorlagen, wurde mittels zusätzlicher Recherche versucht, potenzielle Unternehmensverbindungen zu identifizieren. Vor allem bei den älteren Fällen, wo keine Transparenzberichte vorliegen, kann im Einzelfall aber natürlich nicht ausgeschlossen

Anteil liegt in der Kontrollgruppe lediglich bei 7,9%. Bezieht man nicht nur Erstprüfungen, sondern zusätzlich auch alle Prüfungen mit ein, bei denen der Prüfer aufgrund eines Wechsels im Vorjahr im Fehlerjahr erst das zweite Jahr (in Folge) für den Mandanten die Konzernabschlussprüfung durchgeführt hat, so steigt die entsprechende Quote in der Fehlergruppe auf 23,1% und in der Kontrollgruppe auf 17,7% an. Dabei werden allerdings in jeder Gruppe jeweils vier Fälle aus der Auswertung ausgeschlossen, weil für den jeweiligen Fall oder für seinen Matchpartner in der jeweils anderen Gruppe die notwendigen Prüferdaten nicht verfügbar sind¹⁴³ und im Fehlerjahr nicht bereits schon ein Prüferwechsel vorliegt. Aus analogen Gründen verliert man bei der letzten Auswertung noch einmal zwei Fälle, sodass die auswertbare Stichprobe in der Fehler- und Kontrollgruppe hier jeweils 145 Fälle umfasst. Dabei ergibt sich, dass 31,7% der 145 von der DPR bzw. der BaFin als fehlerhaft eingestuften Konzernabschlüsse bzw. -lageberichte von einem Abschlussprüfer geprüft wurden, der für diesen Konzern erstmals oder zumindest erst das zweite oder dritte Jahr beauftragt war.¹⁴⁴ In der Kontrollgruppe liegt diese Quote bei 24,8%.¹⁴⁵

Für die paarweise Analyse bei expliziter Berücksichtigung des 1-1-Match¹⁴⁶ werden in einem letzten Schritt für jeden der drei in Tab. 2.8 im Hinblick auf einen potenziellen Prüferwechsel ausgewerteten Zeiträume aus allen aus Fehler- und Vergleichsunternehmen gebildeten Paaren jeweils diejenigen herausgefiltert, die hinsichtlich des Auftretens eines Prüferwechsels nicht übereinstimmen.¹⁴⁷ Die Anzahl nicht übereinstimmender Paare kann Tab. 2.9 entnommen werden, wobei danach unterschieden wird, ob der Prüferwechsel (PW) im betrachteten Zeitraum beim

werden, dass eine bestehende Unternehmensverbindung zwischen zwei Prüfungsgesellschaften nicht erkannt und somit ein Prüferwechsel angenommen wird.

¹⁴³ Hierbei handelt es sich um Fälle, für die für das jeweilige Jahr kein Bestätigungsvermerk gefunden wurde und auch aus Thomson Reuters Eikon kein Prüfer entnommen werden konnten.

¹⁴⁴ In den ganz wenigen Fällen, in denen in den betrachteten Jahren bei Unternehmen in der Fehler- oder Kontrollgruppe nicht durchgängig ein Konzernabschluss aufgestellt wurde, wird für die Quotenberechnung (in den Spalten 2, 3 und 4 von Tab. 2.8) jeweils kein Prüferwechsel angenommen, wenn der (die) Einzelabschluss (Einzelabschlüsse) des jeweiligen Unternehmens im (in den) Vorjahr(en) vom gleichen Abschlussprüfer (gem. den Ausführungen in Fn. 142) geprüft wurde(n) wie der nachfolgende Einzel- bzw. Konzernabschluss.

¹⁴⁵ In diesem Zusammenhang ist anzumerken, dass sich unter den vier Unternehmen der Kontrollgruppe, die aufgrund ihres jeweils nicht auswertbaren Matchpartners aus der Analyse ausgeschlossen werden, zusätzlich eines befindet, bei dem ein Prüferwechsel vorgelegen hätte. Bei den zwei Unternehmen der Fehlergruppe, die mangels Matchpartner nicht in die Analyse eingehen, hätte hingegen in keinem Fall ein Prüferwechsel vorgelegen.

¹⁴⁶ Vgl. hierzu Cram/Karan/Stuart, *Contemporary Accounting Research* 2009 insb. S. 479 ff., wo auf die Notwendigkeit hingewiesen wird, für eine korrekte statistische Analyse mit diskreter abhängiger Variable das Matching mittels einer bedingten logistischen Regression explizit zu berücksichtigen. Vgl. zur methodischen Vorgehensweise bei der folgenden Auswertung mittels univariater bedingter logistischer Regression auch Hosmer/Lemeshow/Sturdivant, *Applied Logistic Regression*, 3. Aufl. 2013, Kap. 7.

¹⁴⁷ Hinsichtlich des Auftretens bzw. Nichtauftretens eines Prüferwechsels im jeweiligen Zeitraum übereinstimmenden Paaren kann hingegen kein zusätzlicher Erklärungsgehalt zukommen; vgl. hierzu allgemein Hosmer/Lemeshow/Sturdivant, a.a.O. (Fn. 146), S. 246.

Fehlerunternehmen (Spalte 2) oder beim Vergleichsunternehmen (Spalte 3) des jeweils betrachteten 1-1-Match auftritt.

Tab. 2.8: Prüferwechsel im Jahr des Fehlers oder in den zwei Jahren davor

Fehlergruppe	Im Jahr des Fehlers	Im Jahr des Fehlers oder im Vorjahr	Im Jahr des Fehlers oder in den zwei Jahren davor
Ausgangsstichprobe	151	151	151
abzüglich Fälle mit fehlenden Daten	0	3	4
abzüglich Fälle ohne „Match“	0	1	2
Stichprobe für die Ermittlung der Prüferwechsel	151	147	145
Fälle ohne Prüferwechsel	132	113	99
Fälle mit Prüferwechsel	19	34	46
<i>Quote Prüferwechsel</i>	<i>12,6%</i>	<i>23,1%</i>	<i>31,7%</i>
Kontrollgruppe			
Ausgangsstichprobe	151	151	151
abzüglich Fälle mit fehlenden Daten	0	1	2
abzüglich Fälle ohne „Match“	0	3	4
Stichprobe für die Ermittlung der Prüferwechsel	151	147	145
Fälle ohne Prüferwechsel	139	121	109
Fälle mit Prüferwechsel	12	26	36
<i>Quote Prüferwechsel</i>	<i>7,9%</i>	<i>17,7%</i>	<i>24,8%</i>

Das Verhältnis der in Spalte 2 von Tab. 2.9 eingetragenen Anzahl zur Anzahl von Paaren in Spalte 3 entspricht hinsichtlich des jeweils betrachteten Auswertungszeitraums der sog. odds ratio, die jeweils Spalte 4 zu entnehmen ist. Der Quotient aus der jeweiligen Anzahl in Spalte 2 dividiert durch die Summe der für den jeweiligen Zeitraum insgesamt auftretenden, hinsichtlich eines Prüferwechsels nicht übereinstimmenden Paare (Summe aus Spalte 2 und 3) führt zu der in Spalte 5 angegebenen Wahrscheinlichkeit dafür,¹⁴⁸ dass bei einem hinsichtlich des Prüferwechsels im jeweiligen Zeitraum nicht übereinstimmenden Paar das Unternehmen mit Prüferwechsel ein Unternehmen mit einer Fehlermeldung ist. Die Gegenwahrscheinlichkeit hierzu, dass es sich bei diesem Unternehmen um ein Unternehmen ohne Fehlermeldung handelt, findet sich in Spalte 6. Spalte 7 enthält schließlich den mittels einer univariaten bedingten logistischen Regression geschätzten Koeffizienten,¹⁴⁹ wenn als unabhängige Variable das Auftreten eines Prüferwechsels

¹⁴⁸ Diese entspricht dem Quotienten $\text{odds ratio} / (1 + \text{odds ratio})$.

¹⁴⁹ Dieser entspricht jeweils dem natürlichen Logarithmus der odds ratio. Der p-Wert in Klammern gibt bei Durchführung eines zweiseitigen Tests an, ob der Koeffizient mit 10% (*), 5% (**) oder 1% (***) Irrtumswahrscheinlichkeit signifikant von Null verschieden ist.

im in der jeweiligen Zeile abgegrenzten Zeitraum und als abhängige Variable das Auftreten einer Fehlermeldung verwendet wird.¹⁵⁰

Tab. 2.9: Paarweiser Vergleich von Prüferwechseln im Jahr des Fehlers oder in den zwei Jahren davor

Dichotome Variable D	Anzahl an Paaren mit D = 1 beim Fehlerunternehmen und D = 0 beim Vergleichsunternehmen	Anzahl an Paaren mit D = 0 beim Fehlerunternehmen und D = 1 beim Vergleichsunternehmen	Odds Ratio	Wahrscheinlichkeit, dass bei einem nicht übereinstimmenden Paar ein Unternehmen mit D = 1 das Fehlerunternehmen ist.	Wahrscheinlichkeit, dass bei einem nicht übereinstimmenden Paar ein Unternehmen mit D = 0 das Fehlerunternehmen ist.	Koeffizient der unabhängigen Variable D einer univariaten bedingten logistischen Regression (p-Wert, zweiseitig)
PW im Jahr des Fehlers = 1; kein PW im Jahr des Fehlers = 0 (n = 151 Paare)	18	11	1,64	62,07%	37,93%	0,492 (0,198)
PW im Jahr des Fehlers oder im Vorjahr = 1; kein PW im Jahr des Fehlers und im Vorjahr = 0 (n = 147 Paare)	31	23	1,35	57,41%	42,59%	0,298 (0,278)
PW im Jahr des Fehlers oder in den zwei Jahren davor = 1; kein PW im Jahr des Fehlers und in den zwei Jahren davor = 0 (n = 145 Paare)	40	30	1,33	57,14%	42,86%	0,288 (0,234)

Der letzten Spalte von Tab. 2.9 ist zu entnehmen, dass für keinen der ausgewerteten Zeiträume die aufgezeigten Unterschiede in der Anzahl der Prüferwechsel zwischen Fehler- und Vergleichsunternehmen signifikant sind.¹⁵¹ Insofern liefert diese einfache univariate Analyse keinen Hinweis dafür, dass ein Zusammenhang zwischen tendenziell niedriger Prüfungsqualität

¹⁵⁰ Beide jeweils definiert als dichotome Variable mit Ausprägung Eins im Fall des Auftretens und Null sonst.

¹⁵¹ Da der durchgeführte Matchingprozess, wie in Teil 1, Abschn. 2.3, erläutert, hinsichtlich der Ausprägung des Matchingkriteriums „Unternehmensgröße“ keine exakte Größengleichheit für Fehler- und zugehöriges Vergleichsunternehmen gewährleistet, kontrolliert die in Tab. 2.9 präsentierte univariate Analyse nicht perfekt für den Einfluss der Unternehmensgröße; vgl. hierzu insb. Cram/Karan/Stuart, Contemporary Accounting Research 2009 S. 480. Eine durchgeführte Robustheitsanalyse (vgl. hierzu Cram/Karan/Stuart, Contemporary Accounting Research 2009 S. 494-497), bei der alternativ das Matchingkriterium „Unternehmensgröße“ bzw. die logarithmierte Unternehmensgröße zusätzlich als unabhängige Variable berücksichtigt wird, führt hinsichtlich des Vorzeichens des in Tab. 2.9, Spalte 7, präsentierten Koeffizienten und der Größenordnung des p-Werts zu keinem abweichenden Ergebnis.

und Informationsdefiziten des Prüfers im Hinblick auf den jeweiligen Mandanten aufgrund eines vorausgehenden Prüferwechsels besteht.¹⁵² Ein Argument gegen die Einführung einer externen Rotation ergibt sich hieraus also nicht.¹⁵³

Die hier abgeleiteten Befunde liefern umgekehrt jedoch auch keine Unterstützung für die Einführung einer externen Prüferrotation. Da der Forderung nach einer verpflichtenden externen Rotation regelmäßig die Überzeugung zugrunde liegt, dass sehr lang andauernde Prüfer-Mandanten-Beziehungen mit einer niedrigeren Prüfungsqualität verbunden sind, könnten zukünftige einschlägige Untersuchungen eine Auswertung der bis zum Fehlerjahr jeweils zu beobachtenden Gesamtmandatsdauern bei Fehler- und Vergleichsunternehmen vornehmen. Die hierfür benötigten Daten könnten zumindest teilweise nach dem 17.06.2016 erteilten Bestätigungsvermerken entnommen werden, die gem. Art. 10 Abs. 2 Buchst. b EU-VO neuerdings Angaben zur gesamten bisherigen Mandatsdauer enthalten müssen. Bisher war man für die Feststellung dieser Dauer entweder auf die händische Auswertung aller früheren Bestätigungsvermerke bis zum letzten Prüferwechsel oder auf die entsprechenden Informationen in einschlägigen Datenbanken angewiesen. Beides dürfte regelmäßig dazu geführt haben, dass Mandatsdauern nicht durchgängig für alle Fälle hinreichend weit zurückverfolgt werden konnten.

Überwiegend Daten bis zum Jahr 1998 zurück verwendet in diesem Zusammenhang z.B. die neue länderübergreifende Studie für Europa von *Garcia-Blandon/Argilés-Bosch/Ravenda*, die ebenfalls bereits vor dem Hintergrund der Neuregelung zur externen Rotation durch die EU-VO den

¹⁵² Vergleicht man die Analyseergebnisse mit den in Teil 1, Abschn. 2.2, aufgeführten Studien, die ebenfalls DPR-Fehlermeldungen als Surrogatmaß für Prüfungsqualität verwenden, so zeigt sich, dass die nicht signifikanten Unterschiede hinsichtlich des Prüferwechsels im Fehlerjahr kompatibel sind mit der deskriptiven Analyse bei Gros, *International Journal of Economics and Accounting* 2016 S. 231, dessen Fehlergruppe einen deutlich geringeren, dessen Kontrollgruppe hingegen einen deutlich größeren Umfang als in der vorliegenden Studie aufweist (kein 1-1-Match). Nicht kompatibel sind die Ergebnisse hingegen mit den signifikant häufigeren Prüferwechseln, die bei Frey/Möller/Weinzierl, *KoR* 2016 S. 571 f., insb. Tab. 2.5, für den Dreijahreszeitraum bei ihrer Fehlergruppe abgeleitet werden. Hierbei ist aber z.B. zu berücksichtigen, dass Unternehmen mit eingeschränkten Bestätigungs- oder mit Versagungsvermerken im Gegensatz zur vorliegenden Studie nicht aus der Fehlergruppe ausgeschlossen wurden; vgl. Frey/Möller/Weinzierl, *KoR* 2016 S. 574. Auch die im Rahmen einer Robustheitsanalyse bei Krauß/Pronobis/Zülch, *Journal of Business Economics* 2015 S. 71-76, durchgeführte multivariate logistische Regression zeigt einen schwach signifikant positiven Zusammenhang zwischen einer Erstprüfung und einer DPR-Fehlermeldung auf. Auch hier scheinen die Fälle mit Fehlermeldungen auch die Fälle mit eingeschränkten Bestätigungs- oder mit Versagungsvermerken zu umfassen.

¹⁵³ Dabei ist allerdings zu berücksichtigen, dass für die Untersuchung dieses Abschnitts der Einschränkung eine größere Bedeutung zukommen kann, dass dasselbe Fehler- oder dasselbe Vergleichsunternehmen für unterschiedliche Jahre jeweils mehrfach in der Fehler- bzw. Kontrollgruppe enthalten sein kann. Zur Begründung dieser Vorgehensweise vgl. Teil 1, Abschn. 2.3. Dieses Problem ist in der Kontrollgruppe ausgeprägter als in der Fehlergruppe.

Untersuchungszeitraum 2009 bis 2016 für eine Teilstichprobe des Standard & Poor's Europe 350 betrachtet.¹⁵⁴ Mittels multivariater Analysen und unter Verwendung von drei unterschiedlichen Surrogatmaßen für Prüfungsqualität können dort keine Hinweise dafür gefunden werden, dass bei Unternehmen, die bereits seit mehr als 10 (20) Jahren von demselben Abschlussprüfer geprüft werden, eine signifikant schlechtere Prüfungsqualität vorliegt als bei Unternehmen mit kürzerer bisheriger Gesamtmandatsdauer des Prüfers. Vielmehr ergeben sich bei einigen Analysen sogar Hinweise auf eine höhere Prüfungsqualität, wenn die 10- bzw. 20-Jahres-Schwelle bereits überschritten ist. Die Einführung der externen Pflichtrotation durch die EU-VO wird durch diese Befunde also nicht unterstützt.¹⁵⁵

2.11 Zusammenfassung und Einschränkungen der Untersuchung

Mit dem Ziel der Verbesserung der Qualität der Abschlussprüfung bei PIEs hat die Verordnung (EU) Nr. 537/2014 über spezifische Anforderungen an die Abschlussprüfung bei Unternehmen von öffentlichem Interesse vom April 2014 u.a. die Erbringung von NPL durch den Abschlussprüfer sowohl hinsichtlich der Kategorien zulässiger NPL als auch im Hinblick auf den erlaubten Umfang weiter eingeschränkt. Auch mit der erstmaligen verbindlichen Einführung der externen Prüferrotation soll die Prüfungsqualität verbessert werden.

In Teil 1, Teil 2 und Teil 3 der vorliegenden Studie werden anhand einer deskriptiven Analyse Hinweise zur Beantwortung der Frage gegeben, ob diese Vorgaben der EU-VO sowie die damit

¹⁵⁴ Vgl. Garcia-Blandon/Argilés-Bosch/Ravenda, *Journal of International Financial Management & Accounting*, Version of Record online: 26.04.2019 S. 10. Die Studie enthält auch einen Überblick über weitere empirische Untersuchungen für einzelne Länder Europas, die für die Beurteilung der Einführung der Pflichtrotation durch die EU-VO mit herangezogen werden könnten; vgl. ebenda, S. 4. Vgl. zu einem weiteren nicht auf Europa beschränkten einschlägigen Literaturüberblick z.B. auch Tepalagul/Lin, *Journal of Accounting, Auditing and Finance* 2015 S. 108 ff.

¹⁵⁵ Dabei ist allerdings u.a. die Einschränkung zu beachten, dass das Überschreiten einer Gesamtmandatsdauer von 20 Jahren aus Datenverfügbarkeitsgründen nicht exakt gemessen werden kann; vgl. hierzu Garcia-Blandon/Argilés-Bosch/Ravenda, *Journal of International Financial Management & Accounting*, Version of Record online: 26.04.2019 S. 11 und 27. Als einschlägige deutsche Studie kann z.B. eine ältere, jedoch speziell auf Unternehmen des CDAX im Untersuchungszeitraum 2005 bis 2007 bezogene Studie von Quick/Wiemann angeführt werden, in der mittels multivariater Analysen eine mit der Mandatsdauer ansteigende Prüfungsqualität ableitet wird, wobei diskretionäre Periodenabgrenzungen als Surrogatmaß für Prüfungsqualität herangezogen werden. Dabei ist insb. nach dem dritten Jahr eine verbesserte Prüfungsqualität zu beobachten, wohingegen zwischen mittellangen bisherigen Gesamtmandatsdauern von vier bis maximal sieben Jahren und langen Mandatsdauern von mehr als sieben Jahren kein wesentlicher Qualitätsunterschied mehr abgeleitet wird. Die Mandatsdauern werden in dieser Studie bis 1989 zurück erfasst; vgl. Quick/Wiemann, *ZfB* 2011 S. 923, 929 f. und 933 f. Eine weitere jüngere Studie leitet für deutsche Unternehmen des CDAX einen nicht linearen Zusammenhang zwischen Mandatsdauer und der im vorliegenden Beitrag durchgängig nicht betrachteten (durch die Aktionäre) wahrgenommenen Prüfungsqualität, gemessen durch den Earnings-response-Koeffizienten, ab. Das Maximum der wahrgenommenen Qualität liegt bei 8-9 Jahren Mandatsdauer, wohingegen in den frühen und späteren Jahren der Prüfer-Mandanten-Beziehung die wahrgenommene Qualität geringer ist; vgl. Hohenfels, *International Journal of Auditing* 2016 S. 231.

zusammenhängende Ausübung bzw. Nichtausübung in der EU-VO vorgesehener Mitgliedstaatenwahlrechte im Rahmen des AReG tatsächlich geeignet sind, zur Verbesserung der Abschlussprüfungsqualität beizutragen. Hierzu wird für eine Stichprobe von Fällen mit tendenziell niedriger Prüfungsqualität untersucht, inwieweit die durch die EU-VO vorgesehenen Beschränkungen hinsichtlich Art und/oder Umfang von NPL bei diesen Fällen überhaupt zum Tragen gekommen wären, wenn sie im Untersuchungszeitraum bereits zu beachten gewesen wären. Betrachtet wird schließlich auch, wie häufig bei diesen Fällen mit tendenziell niedriger Prüfungsqualität der zuständige Konzernabschlussprüfer für die Prüfung erst zum ersten bzw. zum zweiten oder zum dritten Mal bestellt war. Als Surrogatmaß für eine niedrige Prüfungsqualität werden dabei zwischen dem 01.07.2005 und dem 31.12.2017 im Bundesanzeiger veröffentlichte Meldungen von Rechnungslegungsfehlern verwendet, die von der DPR bzw. der BaFin im Rahmen eines sog. Enforcementverfahrens festgestellt wurden. Die Bestandsaufnahme wird jeweils mit dem Befund einer Kontrollstichprobe mit tendenziell hoher Prüfungsqualität verglichen.

In Teil 1 der empirischen Bestandsaufnahme wurde ausschließlich die Häufigkeit des Auftretens der im Anhang des Konzernabschlusses gem. § 314 Abs. 1 Nr. 9 HGB zu veröffentlichenden Kategorien von Prüfungs- und Nichtprüfungsleistungen betrachtet. Es ergeben sich keine Hinweise dafür, dass den von der EU als kritisch angesehenen NPL bei den Fällen der Fehlergruppe eine höhere Bedeutung zukommt als in der Kontrollgruppe. Vielmehr sind Unternehmen, bei denen nur Abschlussprüfungsleistungen (APL) oder nach Inkrafttreten des Bilanzrechtsmodernisierungsgesetzes (BilMoG) ausschließlich APL und andere Bestätigungsleistungen (aBL) erbracht werden, eher Unternehmen mit Fehlermeldungen. Steuerberatungsleistungen (StBL) und sonstige Leistungen (sL) sind hingegen deutlich häufiger in der Kontrollgruppe zu beobachten. Insofern ergibt sich hieraus kein Argument für die Verschärfung der bisherigen Regelungen hinsichtlich des Angebots zusätzlicher NPL durch Einführung der black list. Auch sprechen die Ergebnisse der Analyse im Hinblick auf die StBL zumindest nicht gegen die Abschwächung des strikten Verbots, die in dem von der EU in diesem Zusammenhang eingeräumten Mitgliedstaatenwahlrecht zum Ausdruck kommt.

Was die Vorschrift zum cap betrifft, die die Höhe der maximal zulässigen Honorare aus NPL an die Höhe der erzielten Honorare aus APL bindet, so liefert die Auswertung in Teil 2 des Beitrags keine Hinweise dafür, dass der aus EU-Sicht kritische Level von 70% bei Fällen mit tendenziell niedriger Prüfungsqualität häufiger überschritten wird als bei Fällen mit hoher Prüfungsqualität.

Im Übrigen wäre bei 88,9% der in diesem Zusammenhang untersuchten 72 Fälle mit fehlerhaftem Konzernabschluss bzw. -lagebericht die 70%-Honorargrenze gar nicht zum Tragen gekommen. Dies unterstützt die in der Literatur geäußerten Zweifel daran,¹⁵⁶ dass die Einführung des cap in der gewählten Ausgestaltung eine besonders wirksame Maßnahme zur Steigerung der Prüfungsqualität ist.

Auch hinsichtlich der in Art. 4 Abs. 3 EU-VO vorgesehenen 15%-Honorargrenze können durch die Auswertung in Teil 2 Zweifel daran nicht ausgeräumt werden, dass diese EU-Regelung besonders effektiv ist, um breitflächig eine substanzielle Verbesserung der Prüfungsqualität zu erreichen. Das, wenn auch nur über den Gesamtzeitraum, deutlich bessere Ergebnis in der Kontrollgruppe könnte immerhin als Hinweis für die grds. richtige Stoßrichtung der Regelung gedeutet werden. So ergibt die paarweise Analyse von hinsichtlich des Überschreitens der Honorargrenze nicht übereinstimmenden Paaren aus Fehler- und Vergleichsunternehmen, dass die Wahrscheinlichkeit dafür, dass ein Unternehmen, bei dem die 15%-Honorargrenze für das Fehlerjahr überschritten wird, ein Unternehmen mit Fehlermeldung ist, mehr als dreimal so groß ist wie die Wahrscheinlichkeit, dass es sich bei diesem Unternehmen um ein Unternehmen ohne Fehlermeldung handelt. Für die Wirksamkeit der Regelung ist aber offensichtlich deren konkrete Ausgestaltung entscheidend. So erscheint die vorgegebene Grenze von 15% Umsatzanteil angesichts der vorgelegten Bestandsaufnahme als hoch. Aus ökonomischer Sicht erscheint schließlich auch nicht plausibel, dass eine Regelung geeignet sein soll, die Unabhängigkeit des Prüfers zu erhöhen, die auf Prüferseite nur auf den gesetzlichen (Konzern-)Abschlussprüfer abhebt und damit eine Umgehung der Regelung leicht ermöglicht, indem Leistungen für den Mandanten nicht durch den Prüfer selbst, sondern z.B. durch verbundene Unternehmen des Prüfers erbracht werden.¹⁵⁷

Teil 3 des Beitrags hat schließlich ergeben, dass die Unternehmen mit von der DPR bzw. der BaFin als fehlerhaft eingestuften Konzernabschlüssen bzw. -lageberichten häufiger als die Vergleichsunternehmen ohne Fehlermeldungen von einem Abschlussprüfer geprüft werden, der für diesen Konzern erstmals oder aufgrund eines Prüferwechsels in der nahen Vergangenheit zumindest erst das zweite oder dritte Jahr beauftragt ist. Die durchgeführte bedingte logistische

¹⁵⁶ Vgl. z.B. Ratzinger-Sakel/Schönberger, *Accounting in Europe* 2015 S. 62 f. und 78 f., sowie Pott/Schröder/Weckelmann, *KoR* 2014 S. 502.

¹⁵⁷ Vgl. hierzu z.B. bereits die Überlegungen bei Lenz, *DB* 2016 S. 2557 ff. Analog wird dort auch im Hinblick auf die Regelung des Art. 4 Abs. 2 EU-VO argumentiert.

Regression ergibt allerdings, dass die aufgezeigten Unterschiede in der Anzahl der Prüferwechsel zwischen Fehler- und Vergleichsunternehmen für keinen der ausgewerteten Zeiträume signifikant sind. Insofern liefert die univariate Analyse keinen Hinweis dafür, dass ein Zusammenhang zwischen tendenziell niedriger Prüfungsqualität und Informationsdefiziten des Prüfers im Hinblick auf den jeweiligen Mandanten aufgrund eines vorausgehenden Prüferwechsels besteht. Ein Argument gegen die Einführung einer externen Rotation ergibt sich hieraus also nicht.

Wie jede empirische Untersuchung unterliegt auch diese Studie spezifischen Einschränkungen. So muss aufgrund der nicht direkten Beobachtbarkeit von Prüfungsqualität ein Surrogatmaß für schlechte bzw. gute Prüfungsqualität verwendet werden, sodass die Aussagekraft der Studie naturgemäß auch davon abhängt, wie gut das Vorhandensein bzw. Nichtvorhandensein von DPR-Fehlermeldungen niedrige bzw. hohe Prüfungsqualität widerspiegelt. Außerdem werden für die Auswertung die Angaben gem. § 314 Abs. 1 Nr. 9 HGB hinsichtlich der Honorare des Konzernabschlussprüfers herangezogen, deren Aufschlüsselung auf die einzelnen Kategorien ermessensbehaftet ist und bei denen sich die Angabepflichten während des Untersuchungszeitraums durch das BilMoG verändert haben.¹⁵⁸ Einschränkungen der Interpretation der vorliegenden Ergebnisse ergeben sich im Zusammenhang mit den Honoraren insb. auch dadurch, dass die Beachtung der Regelungen der EU-VO teilweise eine Abgrenzung der Abschlussprüferhonorare vorsehen würde, die von der abweicht, die in Deutschland für die Honorarangaben im Anhang der Konzernabschlüsse der PIEs zu beachten ist.¹⁵⁹ Im Übrigen erfolgt die Einschätzung der ausgewählten Regelungen der EU-VO und des AReG ausschließlich vor dem Hintergrund des für den Untersuchungszeitraum in Deutschland geltenden einschlägigen rechtlichen Rahmens. Im Hinblick auf die Beurteilung der Einführung der verpflichtenden externen Prüferrotation ist außerdem zu beachten, dass die vorliegende Studie in dieser Hinsicht schon

¹⁵⁸ Vgl. hierzu die Ausführungen in Teil 1, Abschn. 2.3.

¹⁵⁹ Dies gilt im Übrigen nicht nur in sachlicher Hinsicht für die Abgrenzung des Prüfers und der Leistungsempfänger sowie der Kategorien von APL und NPL (vgl. hierzu insb. die Ausführungen in Teil 1, Abschn. 2.4, und Teil 2, Abschn. 2.7.1), sondern auch in zeitlicher Hinsicht. So wäre für die Berechnung des cap gem. Art. 4 Abs. 2 EU-VO der Durchschnitt der in den letzten drei Jahren gezahlten Honorare für APL relevant; vgl. hierzu z.B. auch CEAOB, Guidelines on „Monitoring the fee cap of non-audit services“ vom 21.09.2018, S. 3, sowie Bose/Lilienbecker, BB 2019 S. 749. Für die Berechnung des 15%-Honoraranteils gem. Art. 4 Abs. 3 EU-VO wäre wiederum relevant, wann die Honorare bei der Prüfungsgesellschaft in die Umsatzerlöse eingehen. Die Angabepflicht der Honorare gem. § 314 Abs. 1 Nr. 9 HGB hebt jedoch weder auf den Zahlungszeitpunkt noch auf den Zeitpunkt der Ertragsrealisation bei der Prüfungsgesellschaft ab; vgl. hierzu z.B. Grottel, in: Grottel u.a. (Hrsg.), Beck Bil-Komm., 11. Aufl. 2018, § 314 HGB Rdn. 161, i.V.m. § 285 HGB Rdn. 503-509.

deshalb keine hinreichende Einschätzung erlaubt, weil die gesamte bisherige Dauer der jeweiligen Prüfer-Mandanten-Beziehung bis zum Fehlerjahr nicht erhoben wurde. Was schließlich den Prozess des 1-1-Matching betrifft, so wird dem trade-off zwischen der Qualität des Match und der Unabhängigkeit der Beobachtungen dadurch Rechnung getragen,¹⁶⁰ dass zwar gleiche Unternehmens-Jahr-Kombinationen nicht in die Kontrollstichprobe aufgenommen werden. Beobachtungen für dieselben Unternehmen in unterschiedlichen Jahren, die möglicherweise ebenfalls nicht völlig unkorreliert sind, werden hingegen nicht aus der Stichprobe ausgeschlossen. Ebenso werden auch Fehlerbekanntmachungen, die sich auf dasselbe Unternehmen, aber auf Konzernabschlüsse bzw. -lageberichte unterschiedlicher Jahre beziehen, in der Fehlergruppe belassen. Schließlich handelt es sich bei dem Beitrag um eine deskriptive Bestandsaufnahme, bei der lediglich univariate Analysen durchgeführt werden. Weitere Studien sollten die Betrachtung der Neuregelungen um multivariate Analysen erweitern,¹⁶¹ wobei ggf. auch alternative Möglichkeiten der Auswahl der Kontrollstichprobe berücksichtigt werden könnten.¹⁶²

¹⁶⁰ Vgl. hierzu allgemein z.B. Shipman/Swanquist/Whited, *The Accounting Review* 2017 S. 217, sowie Stuart, *Statistical Science* 2010 S. 9.

¹⁶¹ Zu multivariaten Analysen hinsichtlich Gesamtmandatsdauer und Prüfungsqualität vgl. bereits die Ausführungen am Ende von Abschn. 2.10. Zu multivariaten Analysen hinsichtlich NPL und Prüfungsqualität vgl. für den deutschen Markt, aber mit von diesem Beitrag jeweils abweichenden Surrogatmaßen für Prüfungsqualität z.B. bereits Hohenfels/Quick, *Review of Managerial Science OnlineFirst* 25.10.2018, sowie Holm, Arbeitspapier 2016, verfügbar bei SSRN: <http://hbfm.link/4839>, insb. S. 21 f. i.V.m. Endn. 7. Hohenfels/Quick untersuchen den Zusammenhang zwischen dem Umfang an NPL (auch aufgeschlüsselt nach Kategorien) und dem Betrag diskretionärer Periodenabgrenzungen. Dieser Zusammenhang ist für NPL, aBL und sL signifikant positiv, für die StBL hingegen insignifikant. Der daraus u.a. abgeleitete negative Einfluss des gesamten Umfangs an NPL auf die Prüfungsqualität bleibt bestehen, wenn Fälle oberhalb des 70%-cap von der Analyse ausgeschlossen werden. Die Autoren leiten einen optimalen cap von etwa 41% ab. Werden nur Fälle unterhalb dieses cap betrachtet, kann gerade kein signifikanter Zusammenhang mehr zwischen den NPL und dem Qualitätsmaß abgeleitet werden. Dies wird als Indiz für eine möglicherweise von der EU zu hoch gewählte prozentuale Grenze interpretiert. Im Gegensatz zu Teil 2, Abschn. 2.7.1, der vorliegenden Studie scheinen Prüferwechsel und nicht durchgängige NPL während des 3-Jahres-Zeitraums jedoch generell nicht berücksichtigt zu werden. Letzteres scheint auch für die Studie von Holm zu gelten, der für Dänemark, Deutschland, Finnland, Schweden und Großbritannien u.a. untersucht, ob bei Unternehmen mit NPL oberhalb des 70%-cap ein stärkerer positiver Zusammenhang zwischen Honoraren für APL und NPL besteht als für Unternehmen mit NPL unterhalb des cap. Der cap wird dabei nur in einer Robustheitsanalyse durchgängig für alle Länder unter Verwendung des Durchschnitts der APL über drei Jahre berechnet. Für Deutschland kann dann kein signifikanter Unterschied zwischen den beiden Unternehmensgruppen mehr abgeleitet werden.

¹⁶² Vgl. zu der Empfehlung Sensitivitätsanalysen mit unterschiedlichen Vorgehensweisen im Hinblick auf das Matching durchzuführen z.B. Caliendo/Kopeinig, *Journal of Economic Surveys* 2008 S. 44 f.

3. The provision of non-audit services under the EU-regulation on the statutory audit of PIEs: evidence based on enforcement announcements in Germany

Abstract: On the one hand, non-audit services (NAS) might pose a threat to auditor independence by increasing the economic bond between auditor and client, on the other hand, the additional knowledge generated by NAS (so called knowledge spillovers) might boost the ability of the auditor to detect misstatements. This study addresses the question of whether the limitations on NAS introduced by the Regulation (EU) No 537/2014, which applies to the statutory audit of public interest entities (PIEs), might improve audit quality. The analysis is done on a sample of German listed companies subject to enforcement findings from July 2005 until 31 December 2017 and a control group. The results suggest that certain NAS, like tax services, might improve audit quality, which is why there seem to be no need of further restricting the provision of NAS in Germany. Also, the results do not show that there is an association between exceeding the 70% cap on NAS fees and audit quality or that audit quality is reduced when the auditor obtains more than 15% of his revenues from a client. Therefore, the analysis does not provide evidence on whether these limitations are useful measures to improve audit quality.

I appreciate the helpful suggestions and comments of Renate Hecker who motivated me for this paper. It directly builds on and complements the descriptive and univariate analyses in Fiallo Hecker (2019a), (2019b) and (2019c) and benefits from most of the data, numerous analyses, thoughts and results of these three former publications as well as from the fruitful discussions we had while writing these publications.

3.1 Introduction

Whether statutory auditors should provide non-audit services (NAS) to their clients is a much-debated topic in accounting research. On the one hand, NAS might pose a threat to auditor independence by increasing the economic bond between auditor and client¹, on the other hand, the additional knowledge generated by NAS (so called knowledge spillovers) might boost the ability of the auditor to detect misstatements². It follows that, at least in theory, the provision of NAS might both improve and worsen audit quality, which raises the question of how the matter should be regulated. Even though the auditing theory and the empirical results have not solved this dilemma, the European Union has issued in April 2014 Directive 2014/56/EU, which amends the existing Directive 2006/43/EC, and Regulation (EU) No 537/2014³. The EU-regulation, which applies to the statutory audit of public interest entities (PIEs⁴), aims at improving auditor independence and audit quality by limiting the provision of certain NAS (black list) and by setting a cap on NAS fees as well as on the proportion of fees an auditor can obtain from a client⁵.

In this paper, I analyse whether the limitations on NAS introduced by the EU-regulation might improve audit quality⁶. The consequences of the provision of NAS have been discussed extensively in previous studies, so far with mixed results⁷. Prior studies mostly measure audit quality with auditors' going concern/modified opinions, discretionary accruals and abnormal returns⁸. In this paper, audit quality is measured with the existence of accounting errors in audited financial statements, without the auditor having issued a qualified or adverse opinion. Companies with low audit quality are identified by searching for the enforcement announcements, where errors, found

¹ Ostrowski/Söder (1999), p. 562; DeAngelo (1981a), p. 113. On the interaction between auditor independence (or lack thereof) and audit quality, see also Fiallo/Hecker (2019a), p. 226.

² Simunic (1984), pp. 680-681. On the potential benefits of knowledge spillovers on audit quality, see Fiallo/Hecker (2019a), p. 231. A more detailed analysis of the theoretical arguments beyond knowledge spillovers is included in section 3.3.1.2 of this study.

³ Hereafter called EU-regulation or EU-reg.

⁴ The term PIE refers to capital market-oriented companies that have already been admitted to trading in an EU state, as well as certain banks and insurance companies. For a more detailed definition of PIE as well as references to the regulatory framework, see e.g. Fiallo/Hecker (2019a), p. 225 footnote 2.

⁵ Recital 7 and 8 EU-reg.

⁶ This is also the research question analysed by Fiallo/Hecker (2019a), p. 225.

⁷ E.g. see the reviews of prior studies by Ratzinger-Sakel/Schönberger (2015) on the interaction between the provision of NAS and auditor independence, or by Velte/Loy (2018) on the impact of NAS on earnings quality, audit quality and investor perceptions. Also, the results of prior German studies on NAS are reviewed in section 3.3.2.

⁸ Simnett/Carson/Vanstraelen (2016), p. 14.

during the examination of a company's financial statements, are disclosed⁹. The research sample includes German listed companies subject to enforcement findings from July 2005 until 31 December 2017 and a control group¹⁰.

This study is conducted in Germany, where data on erroneous accounting are available thanks to the transparent disclosure of enforcement findings. Also, the Member State options available in the EU-regulation call for country-specific evidence on the necessity (or lack thereof) of making the provisions more/less restrictive at national level. Besides the already mentioned study of *Fiallo/Hecker* (2019a, 2019b and 2019c), I am aware of only few other studies on NAS fees, which use enforcement announcements to measure audit quality: e.g. that of *Markelevich/Rosner* (2013) set in the US and that of *Gros* (2016) set in Germany¹¹. *Gros* (2016) analyses a sample of companies with erroneous financial statements, covering financial years between 2005 and 2012¹². I extend his work by analysing all announcements released up to 31 December 2017 and by focusing on the limitations on NAS introduced by the EU-regulation.

This paper tries to answer three questions. First, as the EU regulator has expressed concerns that the provision of certain NAS compromises auditor independence¹³, I analyse if and how certain types of services provided by the auditor affect audit quality. Second, as the level of fees and their structure are also suspected of impairing auditor independence¹⁴, I analyse if exceeding the cap on NAS fees introduced by the EU-regulation influences audit quality. Third, I try to address the threat identified by the EU regulator of an auditor becoming excessively dependent on a client¹⁵, by testing whether audit quality decreases when the auditor obtains more than 15% of his revenues from a client.

⁹ For an analysis of the arguments in favour and against the use of this surrogate, see sec. 3.4.1 as well as *Fiallo/Hecker* (2019a), p. 226.

¹⁰ This is essentially the sample used in *Fiallo/Hecker* (2019a), p. 228. For more details on sample selection, see sec. 3.4.2.

¹¹ Additionally, *Krauß/Pronobis/Zülch* (2015), pp. 71-75, who analyse the effect of abnormal audit fees on audit quality, use enforcement announcements in their robustness analysis to proxy for audit quality.

¹² Additionally, *Gros* (2016) pp. 219-220, possesses data on enforcement examinations, which ended without enforcement findings. These data are not available in the public domain and my request to the Federal Financial Supervisory Authority (BaFin) to obtain such data for research purposes was denied.

¹³ Recital 8 EU-reg.

¹⁴ Recital 7 EU-reg.

¹⁵ Recital 7 EU-reg.

The results do not provide evidence that NAS increase economic bonding. On the contrary, they show that certain NAS, like tax services, might even improve audit quality. In this case, extending the already existing restrictions on the provision of NAS does not seem necessary¹⁶, while the decision of the German regulator to allow some tax services seems appropriate¹⁷. There is no evidence that the introduction of the cap is a useful measure to improve auditor independence and audit quality. This is in line with prior studies, which have questioned the ability of the cap to have a significant impact on the audited entities¹⁸. Finally, there is no indication that audit quality is reduced when the auditor obtains more than 15% of his revenues from a client¹⁹. However, as theoretical considerations exist which indicate that client importance might impair auditor independence²⁰, it might be wise to limit the financial dependence of auditors on a client.

This paper contributes to the literature in two ways. First, by analysing the limitations introduced by the EU-regulation regarding the provision of NAS and the level of fees charged by the auditor, it provides empirical evidence on the potential impact of certain aspects of the EU-regulation. The results are of interest to the European regulator and, given the options available to Member States, also to the national regulator when assessing the necessity (or lack thereof) of adapting the provisions. Second, this paper adds to the existing literature by using a different proxy for audit quality, namely the existence of enforcement findings. Measuring audit quality has always been a challenge for accounting researchers. The main issue is that audit quality is costly to evaluate, because the procedures performed by the auditor are not directly observable (competence and execution) and because the specific client-auditor dynamic is unknown to outside parties (independence)²¹. If audit quality is visualised as a continuum, at the lowest point there is audit failure, which occurs when the auditor issues an incorrect audit opinion due to either lack of

¹⁶ A similar interpretation of the results is offered in Fiallo/Hecker (2019a), p. 234.

¹⁷ A similar interpretation of the results is provided by Fiallo/Hecker (2019a), p. 234 and by Hohenfels/Quick (2018), OnlineFirst 25 October 2018. Available at: <https://link.springer.com/article/10.1007/s11846-018-0306-z>.

¹⁸ E.g. Fiallo/Hecker (2019c), p. 347; Ratzinger-Sakel/Schönberger (2015), p. 62; Pott/Schröder/Weckelmann (2014), p. 502. Other researches also suggest that the 70% threshold might be too high, see e.g. Hohenfels/Quick (2018), OnlineFirst 25 October 2018. Available at: <https://link.springer.com/article/10.1007/s11846-018-0306-z>.

¹⁹ Conversely, the descriptive analysis of Fiallo/Hecker (2019b), p. 296, shows weak evidence that companies with a fee-to-revenue ratio higher than 15% are more likely to have enforcement findings.

²⁰ DeAngelo (1981b), pp. 187-192. The importance of a client for the auditor can be approximated by the fees paid by that client in relation to the auditor's total revenue. In this study, I define client importance as a fee-to-revenue-ratio higher than 15%, following the threshold set in art. 4 para. 3 EU-reg. It follows that a client is important when he contributes to more than 15% of the auditor's revenue.

²¹ DeAngelo (1981b), p. 186. Similar considerations are made in Fiallo/Hecker (2019a), p. 226.

independence, or the performance of audit procedures that are not in accordance with the prescribed auditing standards²². According to this definition, the presence of material errors in audited financial statements and the lack of a modified audit opinion can be interpreted as an audit failure, and a sign of poor audit quality²³. Since enforcement findings are a concrete and reliable measure of accounting errors, which are found during an official enforcement examination²⁴, they are suitable proxies for low audit quality.

This paper is organised as follows. Section 3.2 describes and outlines the institutional framework focusing on the limitations on NAS introduced by the EU-regulation. In section 3.3, I present the economic bonding and knowledge spillover theories, review prior studies and formulate the research hypotheses. In section 3.4, I describe the research design and in section 3.5m I present the results of the research. The conclusion is outlined in section 3.6.

3.2 Institutional framework

On 17 June 2016, the EU-Directive 2014/56/EU and the EU-regulation 537/2014 came into force. While the EU-regulation finds direct application, the Directive and the Member State options have been implemented via the German Audit Reform Act (Abschlussprüfungsreformgesetz – AReG). The EU-regulation has introduced limitations to the provision of certain NAS and to the level of fees an auditor can receive from a client. These provisions, which apply to the auditors of PIEs, are included in art. 4 and 5 of the EU-reg. and are briefly discussed in this section²⁵.

Art. 5 EU-reg. lists which services the auditor of a PIE and the members of his network cannot provide to the audited entity, its parent and its controlled undertakings, from the beginning of the audited period until the issuing of the auditor's report²⁶. The ban affects at least partially the following categories: tax services, services that involve playing any part in the management or in

²² Francis (2004), p. 346; Francis (2011), pp. 127-129; Markelevich/Rosner (2013), p. 1592.

²³ However, as the auditor attests with reasonable assurance that financial statements are free from material errors, the presence of errors does not always imply an audit failure. This issue as well as other limitations to this choice of proxy are outlined in sec. 3.4.1. Similar considerations are also made in Fiallo/Hecker (2019a), pp. 226-227.

²⁴ Tritschler (2013), p. 144.

²⁵ In this section I provide an overview of the content of art. 4 and 5 EU-reg. For a more detailed analysis of certain aspects of the EU-regulation, including the comparison with the (prior) national regulation, see Fiallo/Hecker (2019a), p. 230 and Fiallo/Hecker (2019b), pp. 290-291, 294.

²⁶ Services related to the establishment of internal controls for the preparation/control of financial information as well as services involving the design and implementation of financial information technology systems, cannot be provided even in the financial year preceding the audited period.

the managerial decisions of the audited company, bookkeeping services, preparation of financial statements, payroll services, services related to the establishment of internal controls for the preparation/ control of financial information, services involving the design and implementation of financial information technology systems, valuation services, legal services, internal audit services, human resources services, and services linked to the financing and investment strategy of the audited entity and linked to the promotion of or dealing with shares in the audited entity. Art. 5 paras. 2 and 3 EU-reg. allow Member States to either extend the list of prohibited services or to make it less restrictive. This last option, which the German regulator has implemented²⁷, regards certain tax and valuation services, which can be provided, if they have no direct and material effect on the audited financial statements and do not compromise the auditor's independence²⁸.

Art. 4 para. 2 EU-reg. introduces a cap on the fees charged for NAS other than those prohibited by art. 5 EU-reg. When the auditor provides NAS to the audited entity, its parent or controlled undertakings for three consecutive years, the amount of NAS fees in the fourth year cannot exceed 70% of the average audit fees paid by the entity over the past three years. The audit fees include those paid for the statutory audit of any parent and controlled undertakings of the audited entity as well as for the group audit. In contrast to the provisions included in art. 5 EU-reg., only the fees paid to the auditor are relevant, those paid to other members of the group or the network are disregarded by the calculation of the cap. Art. 4 para. 2 EU-reg. also leaves it to Member States to regulate the possibility, in exceptional cases, to waive the application of the cap for a maximum period of two years. The German regulator has decided that the Auditor Oversight Body (AOB)²⁹ can allow NAS fees to exceed the cap for one year, given that they do not exceed 140% of the average audit fees³⁰.

Finally, art. 4 para. 3 EU-reg. sets limitations to the total fees an auditor can obtain from one client. Only the fees paid by the PIE to the auditor are relevant here. Fees paid to members of the group or the network are disregarded. If total fees exceed 15% of an auditor's total revenues, for more than three consecutive years, the auditor must disclose the matter to the audit committee, which assesses the possible threat to auditor independence and determines whether there is a need to

²⁷ Sec. 319a para. 1 no. 2-3 German Commercial Code (HGB).

²⁸ Art. 5 para. 4 EU-reg. states in fact that an auditor can provide any non-restricted NAS only with the approval of the audit committee, which must assess any possible threat to independence. Additionally, the effects of the provision of such services must be documented and explained in an additional report.

²⁹ In German: Abschlussprüferaufsichtsstelle (APAS).

³⁰ Sec. 319a para. 1a HGB.

appoint another auditor to perform an engagement quality control review. Furthermore, it is the duty of the audit committee to decide, in the case that fees continue to exceed 15%, if the auditor can carry out the audit for an additional period of maximum two years.

3.3 Theoretical background, literature review and hypothesis development

3.3.1 Theoretical background

3.3.1.1 Audit quality and auditor independence

The purpose of statutory audits is to verify that financial statements are a fair presentation of a company's financial situation and that they are prepared in accordance with the applicable accounting standards³¹. By performing the necessary audit procedures, the auditor attests with reasonable assurance that financial statements are free from material errors³². As the auditor's work is fundamental for the credibility of financial information and the functioning of the capital market, it must be carried out with sufficient audit quality³³.

According to the definition of *DeAngelo* (1981a), audit quality is the auditor's perceived ability to detect misstatements and to report them³⁴. To be able to detect misstatements in the first place, the auditor must be competent, which means possessing the necessary technical capabilities and executing the correct audit procedures³⁵. The second determinant of audit quality is auditor independence or in other words his willingness to report any material misstatements. Being independent means to be free from personal and economic influences that can compromise professional judgement, integrity and objectivity³⁶. This aspect, called independence of mind or independence in fact, is necessary but not sufficient, as the auditor must also be perceived as independent. To achieve independence in appearance the auditor should avoid facts and conditions, which can make him appear less than perfectly independent in the eyes of an informed third party³⁷.

³¹ ISA 200.3

³² ISA 240.5

³³ See also the similar considerations made in Fiallo/Hecker (2019a), pp. 226-227.

³⁴ DeAngelo (1981b), p. 186. This definition provided by DeAngelo is commonly used in auditing literature, see e.g. Fiallo/Hecker (2019a), p. 226.

³⁵ Köhler/Marten/Meuwissen/Quick (2013), p. 242.

³⁶ IESBA (2018), sec. 120.12 A1.

³⁷ IESBA (2018), sec. 120.12 A1.

The provision of NAS might threaten independence in several ways³⁸. First, the advisory activity frequently involves an intensive collaboration between the auditor and the company's representatives. This might lead the auditor to develop a trusting relationship with the client (familiarity threat)³⁹. Also, the provision of services that involve promoting or defending a client's position, might lead the auditor to identify with the client (advocacy threat). In these cases, the provision of NAS might contribute to social bonding⁴⁰, which in turn would reduce the auditor's objectivity and professional scepticism. Second, the more extensive the services the auditor provides, the higher the probability that he will have to assess his own work (self-review threat). In this case, the auditor will lack objectivity. Third, the provision of NAS may increase the financial interest the auditor has in the client (self-interest threat), which might lead to impaired independence, especially if the client threatens him with the termination of the contract (intimidation threat).

The impact each threat has on auditor independence depends on the existence of safeguards⁴¹. For example, the existing limitations on certain NAS reduces the possibility of self-review. Due to the organisational structure of many audit firms, audit services and NAS are often provided by different teams, which might decrease the risk of social bonding. Regarding economic bonding, the regulator seems to think that limiting the provision of certain NAS and the amount of fees an auditor can obtain from a client might improve independence by reducing the self-interest threat⁴². The next section presents the conceptual framework behind economic bonding on the base of the quasi-rents model developed by *DeAngelo* (1981a).

3.3.1.2 Economic bonding and knowledge spillovers

In a world where the audit market is perfectly competitive due to all auditors having the same technological capabilities, the incumbent auditor has a cost advantage compared to other potential

³⁸ According to IESBA (2018), sec. 120.6 A3, the threats to independence can be classified in five categories: self-interest threat, self-review threat, advocacy threat, familiarity threat and intimidation threat.

³⁹ Typically, the familiarity threat arises when the auditor and the client work together for many years, which is an argument in favour of limiting the duration of audit engagements, see e.g. Fiallo/Hecker (2019c), p. 344.

⁴⁰ Hohenfels/Quick (2018), OnlineFirst 25 October 2018. Available at: <https://link.springer.com/article/10.1007/s11846-018-0306-z>.

⁴¹ See also the similar considerations made by Tebben (2011), pp. 224-225.

⁴² Recital 7 and 8 EU-reg. In this study, I focus on the issue of economic bonding. To what extent limiting the provision of NAS also helps in reducing social bonding remains unclear, due to the difficulties in disentangling the consequences of economic and social bonding. According to recital 21 EU-reg., limiting the duration of the audit engagement should reduce the familiarity threat and the risk of social bonding. These aspects are further discussed in Fiallo/Hecker (2019c), pp. 344-347.

auditors due to the start-up costs a new auditor would face and to the transaction costs, the audited company would sustain when switching auditor⁴³. The incumbent auditor is therefore able to set the audit fees above the audit costs, being able to obtain client-specific quasi-rents from future audits.

The existence of quasi-rents has two consequences: first, to win the client the auditor is willing to reduce the fees for the initial audit below the actual costs, so called low-balling⁴⁴; second, the auditor needs to be reappointed to earn the quasi-rents. While the costs related to low-balling are sunk and do not affect the auditor's decisions, the existence of quasi-rents and the willingness to retain them might impair auditor independence. Under these circumstances, the auditor's decisions might be affected by his fear of dismissal. Usually, managers cannot directly threaten the auditor with the termination of the audit engagement (e.g. in German public companies, the auditor is chosen by the shareholders' meeting on a proposal of the supervisory board⁴⁵). However, they might be able to influence the appointment, e.g. if asked to express an opinion on the work of the auditor, or in those cases where the interests of management and supervisory board are aligned⁴⁶. Moreover, managers might exert pressure on the auditor by withdrawing contracts for NAS⁴⁷.

Since auditors usually have more than one client, the risk posed by client-specific quasi-rents must be evaluated considering also the quasi-rents generated by other clients⁴⁸. If an auditor fails to report a misstatement and this information becomes public, he might suffer damage to his reputation, which could lead to the loss of other clients⁴⁹. The auditor's reporting decision results therefore from the cost-benefit consideration of keeping a client against the probability of being discovered "cheating" and losing other clients. It follows that even though quasi-rents of a specific client are a threat to independence, quasi-rents of other clients are "a collateral against opportunistic behaviour"⁵⁰.

⁴³ DeAngelo (1981a), pp. 118-123.

⁴⁴ The auditor is willing to reduce the fees until the point where the net present value of the engagement becomes zero, see DeAngelo (1981a), p. 122.

⁴⁵ Sec. 318 para. 1 sentence 1 HGB and sec. 124 para. 3 AktG.

⁴⁶ Lopatta/Kaspereit/Canitz/Maas (2015), p. 565; Qandil (2014), pp. 95-97.

⁴⁷ Qandil (2014), pp. 95-97.

⁴⁸ DeAngelo (1981a), p. 117; DeAngelo (1981b), p. 184. Similar considerations are included in Fiallo/Hecker (2019b), p. 296.

⁴⁹ Evidence on the existence of reputation concerns, which counteract the economic interest, have been found by Larcker/Richardson (2004), p. 655.

⁵⁰ DeAngelo (1981b), p. 197.

The quasi-rents model has subsequently been extended with the inclusion of a market for NAS⁵¹. There might be several advantages related to the simultaneous provision of audit and non-audit services⁵². As client and auditor have already established a relation, transaction costs related to the purchase of NAS may be lower for both parties. Further, the provision of NAS might generate economies of scope as the auditor can transfer knowledge among audit and non-audit services. It follows that knowledge spillovers can make the joint provision of audit and non-audit services cheaper than their separate provision. The additional knowledge gained through the provision of NAS can generate an information advantage, which might improve the auditor's ability to detect errors thus increasing audit quality⁵³.

Prior studies have looked for evidence of knowledge spillovers. *Simunic* (1984) finds that auditors providing additional NAS are paid higher audit fees compared to auditors, who do not render any NAS, which might indicate that the cost functions of audit and non-audit services are interdependent due to knowledge spillovers⁵⁴. These findings are only partially replicated by *Palmrose* (1986), who finds a positive relation between audit and non-audit fees also when NAS are not provided by the incumbent auditor⁵⁵. Any evidence of the existence of knowledge spillovers indicates that there are cost benefits due to the joint provision of NAS and audit services. However, it remains unclear if this positively affects audit quality, as the cost savings might increase the quasi-rents due to the additional cost advantage the incumbent auditor has from NAS-related start-up and switching costs⁵⁶.

Beck/Frecka/Solomon (1988) argue that non-recurring NAS increase the economic bond, as cost savings increase the value of future rents, while recurring NAS may decrease the incremental bonding and, when the cost savings in audit start-up costs exceed NAS start-up and switching costs, even the total bonding⁵⁷. These results are criticised by *Ostrowski/Söder* (1999), who argue that under more realistic assumptions the provision of NAS always impairs auditor independence⁵⁸.

⁵¹ E.g. Beck/Frecka/Solomon (1988); Simunic (1984); Ostrowski/Söder (1999).

⁵² For considerations made in this regard, see e.g. Simunic (1984), p. 680; Zwernemann (2015), pp. 39-50; Sattler (2011), pp. 67-76; Arruñada (1999), pp. 513-515; Lange (1994), pp. 28-32.

⁵³ Knowledge might be transferred also from audit services to NAS. For evidence that knowledge spillovers move in this direction, see Antle/Gordon/Narayanamoorthy/Zhou (2006), p. 258.

⁵⁴ Simunic (1984), pp. 681, 689.

⁵⁵ Palmrose (1986), p. 406.

⁵⁶ Beck/Frecka/Solomon (1988), pp. 57-61; Ostrowski/Söder (1999), pp. 559-560.

⁵⁷ Beck/Frecka/Solomon (1988), pp. 57-61.

⁵⁸ Ostrowski/Söder (1999), p. 561.

Arruñada (1999) suggests putting the client-specific quasi-rents in relation to all quasi-rents from other clients, which are also higher due to the provision of NAS⁵⁹. He concludes that the risk to independence should be evaluated by considering all the auditor's clients, and that client diversification prevents opportunistic behaviour.

In conclusion, the provision of NAS might generate knowledge spillovers, which would make the joint provision of audit and non-audit services convenient. The extent to which this is beneficial to audit quality remains unclear. On the one hand, the additional knowledge and information gained through the provision of NAS might increase the ability of the auditor to detect errors⁶⁰, on the other hand, the provision of NAS might generate additional quasi-rents, also due to the cost savings related to knowledge spillovers, which might strengthen the economic bond and lead to impaired independence.

3.3.2 Prior literature

3.3.2.1 The provision of NAS and independence in appearance

There are many studies focusing on the question of whether NAS impair auditor independence of mind and in appearance⁶¹. In this and the next sections, I present a selection of prior research conducted in Germany, as the results of international studies are less likely to apply to the German audit environment, due to different institutional characteristics, such as auditor's liability, accounting regime, ownership structure and NAS-related characteristics (e.g. admissible NAS)⁶².

Prior studies, which conduct surveys among e.g. auditors, supervisory board members, shareholders, business journalists, credit and/or financial analysts, mostly report mixed results regarding the relation between NAS and independence in appearance⁶³. According to the experimental analysis of *Meuwissen/Quick* (2009), supervisory board members of companies listed on the DAX, MDAX, SDAX and TecDAX perceive the provision of NAS as negative⁶⁴. Human

⁵⁹ Arruñada (1999), p. 520.

⁶⁰ As already mentioned in Fiallo/Hecker (2019a), pp. 226; DeAngelo (1981a), p. 116, points out that the ability of the auditor to detect errors is also conditional to his willingness to report them.

⁶¹ E.g. see the literature reviews by Ratzinger-Sakel/Schönberger (2015) on the interaction between the provision of NAS and auditor independence, or by Velte/Loy (2018) on the impact of NAS on earnings quality, audit quality and investor perceptions. For a definition of independence of mind and independence in appearance, see sec. 3.3.1.1.

⁶² Tebben (2011), p. 4; Sattler (2011), p. 210.

⁶³ See the review included in *Meuwissen/Quick* (2009), pp. 389-394 and van Liempd/Quick/Warming-Rasmussen (2019), pp. 5-6.

⁶⁴ *Meuwissen/Quick* (2009), pp. 398-401.

resources services are regarded as highly critical for auditor independence (61% of the respondents), followed by services in the area of financial information systems (48%) and by tax services (38%).

Quick/Warming-Rasmussen (2009, 2015) find that according to German private investors the provision of NAS enhances the risk of economic bonding, which is why high levels of NAS fees are perceived as critical⁶⁵. The concerns of investors extend to all types of services, indicating that they believe that the consequences of economic bonding offset any knowledge spillovers. Conversely, a recent survey reported by *Rossmannith/Bleicher* (2017) shows that company's representatives perceive the joint provision of audit and non-audit services as positive, as it improves the quality of financial statements⁶⁶.

Finally, *Eilifsen/Quick/Schmidt/Umlauf* (2018) finds that investors' perceptions have changed over time⁶⁷. While during the period 2005-2007 and during the financial crisis (2008-2009) investors perceived large fees negatively, in the years 2010 until 2015 they show a positive attitude towards large NAS fees. The authors argue that even though the cause of this change is not entirely clear, it is possible that investors might believe that the lesson from the crisis has been learned and that supervisory boards and auditors are now more closely monitoring the provision of NAS⁶⁸. This explanation seems to be supported by the evidence they collect regarding a drop in the provision of NAS following the financial crisis.

3.3.2.2 The provision of NAS and independence of mind

Other studies focus on independence of mind. The analysis of *Tebben* (2011) focuses on companies listed between 2005-2007 on the DAX, MDAX, SDAX and TecDAX and shows that absolute discretionary accruals are positively associated with NAS fees, which seems to indicate the existence of economic bonding⁶⁹. Additionally, he analyses the direction of discretionary accruals and finds that only negative discretionary accruals are positively associated with NAS fees, which means that auditors who receive high NAS fees might be more lenient towards income-decreasing

⁶⁵ Quick/Warming-Rasmussen (2009), pp. 155-156. Quick/Warming-Rasmussen (2015), pp. 164-171.

⁶⁶ The survey was conducted by Albstadt-Sigmaringen University in cooperation with the WirtschaftsTreuhand GmbH. The results are reported by Rossmannith/Bleicher (2017), pp. 227-230.

⁶⁷ Eilifsen/Quick/Schmidt/Umlauf (2018), pp. 298-299.

⁶⁸ Eilifsen/Quick/Schmidt/Umlauf (2018), pp. 311-312.

⁶⁹ Tebben (2011), pp. 135-142.

earnings management practices. Similarly, *Molls* (2013), using a sample of German companies listed on the regulated market between 2005 and 2010, finds evidence of a positive association between NAS fees and negative discretionary accruals⁷⁰.

Hohenfels/Quick (2018) find for a sample of German companies listed on the CDAX between 2006 and 2013 a negative relation between NAS fees and audit quality, measured with discretionary accruals⁷¹. Their results might indicate that the economic and social bonding derived by the provision of NAS exceeds any positive externalities due to knowledge spillovers.

Krauβ/Pronobis/Zülch (2015), whose main analysis focuses on the relation between abnormal audit fees and audit quality, find, for a sample of German companies listed on the regulated market between 2005 and 2010, that abnormal NAS fees are not significantly associated with discretionary accruals⁷². Also, *Zimmermann* (2008) finds no significant association between NAS fees and discretionary accruals, for a sample of German companies listed in 2005 on the DAX, TecDAX, SDAX, MDAX und GEX⁷³.

The analysis of *Ratzinger-Sakel* (2013) shows, for a sample of German financially distressed companies during 2005-2009, that there is no significant negative relation between NAS fees and the likelihood of issuing a going concern opinion (GCO), which might indicate that market-based mechanisms related to reputation concerns dominate the economic interest⁷⁴. Additionally, she analyses whether the joint effect of NAS fees and auditor size affects audit quality. The results show that Big 4 auditors are less likely to issue a GCO in the presence of high NAS fees and highly distressed clients, which could indicate that the relative importance of NAS contracts for Big 4 auditors affects their independence⁷⁵.

Finally, *Dobler* (2014) finds evidence of knowledge spillovers for auditors of Germany's largest private family firms⁷⁶. His results show that, after controlling for joint determination, there is still a positive association between NAS fees and audit fees, which is consistent with interdependencies

⁷⁰ Molls (2013), pp. 225-227.

⁷¹ Hohenfels/Quick (2018), OnlineFirst 25 October 2018. Available at: <https://link.springer.com/article/10.1007/s11846-018-0306-z>.

⁷² Krauß/Pronobis/Zülch (2015), p. 77.

⁷³ Zimmermann (2008), p. 200.

⁷⁴ Ratzinger-Sakel (2013), p.130.

⁷⁵ Ratzinger-Sakel (2013), pp.144-150.

⁷⁶ Dobler (2014), p. 442.

of the cost functions of audit and non-audit services and the existence of knowledge spillovers. However, the author does not interpret his findings in terms of audit quality.

3.3.2.3 Future fees and audit quality

Few studies have analysed the effects of future fees on audit quality⁷⁷. *Blay/Geiger* (2013) and *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018) find that audit quality is negatively affected by the expectation of future fees⁷⁸. Conversely, *Fuhrmann* (2016) finds no evidence of an association between future NAS fees and audit quality⁷⁹. Other studies focus on future NAS growth. *Causholli/Chambers/Payne* (2014) provide evidence that clients, with relatively low NAS in the current period and an increase in the purchase of NAS in the following period, have larger discretionary accruals in the current period. This might indicate that the opportunity for selling additional NAS in the future impairs current auditor independence, thus lowering audit quality. Conversely, *Cahan/Emanuel/Hay/Wong* (2008) and *Sattler* (2011) do not find any significant association between earnings management and future NAS growth⁸⁰.

3.3.2.4 NAS categories and audit quality

Some studies additionally focus on the type of service, as audit quality could be affected not only by the level of fees, but also by the nature of the services rendered. According to sec. 314 para. 1 no. 9 HGB, companies must disclose the fees paid each year to their group auditor according to four categories, namely fees for: a) statutory audit of financial statements, b) other assurance services, c) tax services and d) other non-audit services.

The study of *Fiallo/Hecker* (2019a) shows, for a sample of German companies falling under the enforcement regime, that companies buying only audit services are more likely to have enforcement findings, while companies buying tax services and other non-audit services are less likely to have enforcement findings⁸¹. It follows that the study does not provide indications of the

⁷⁷ On the interaction between future fees and audit quality, see chapter 4 as well as the more extensive review of prior studies outlined in section 4.2.2.

⁷⁸ *Blay/Geiger* (2013), pp. 580-581; *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018), pp. 4-6.

⁷⁹ *Fuhrmann* (2016), p. 57.

⁸⁰ *Cahan/Emanuel/Hay/Wong* (2008), p. 182; *Sattler* (2011), pp. 409-412.

⁸¹ *Fiallo/Hecker* (2019a), pp. 233-234. The authors run a univariate conditional logistic regression. The coefficients of the variable of interest are significant at least at the 5% level. As a robustness check, company's size, measured respectively as total assets and the natural logarithm of total assets, is added as a control variable. The coefficient of the main variable maintains the same sign and significance.

necessity of further restricting NAS. Rather the findings seem to support the decision of the German regulator to allow some tax and valuation services.

For a sample of major listed companies between 2004 and 2011, *Krauß/Zülch* (2013) find that the provision of NAS, particularly of other assurance services, is negatively associated with audit quality, measured with discretionary accruals⁸². They argue that the economic importance of other assurance services, in terms of fees, might tighten the economic bond and outweigh any positive externalities related to knowledge spillovers⁸³. Regarding tax services, their results do not show any significant association with audit quality. *Krauß/Zülch* (2013) provide two different interpretations: it is either possible that the economic bonding and the knowledge spillover effects offset each other, or that, due to the prohibition of many tax services and their consequent low economic importance, both effects are marginal⁸⁴. Finally, they find weak evidence that other non-audit services are negatively associated with audit quality. Even though many other non-audit services are prohibited, they might still generate significant fees, which is why economic bonding could exceed any positive externalities related to knowledge spillovers⁸⁵. Similar results are obtained by *Hohenfels/Quick* (2018), who find, for a sample of German companies listed on the CDAX between 2006 and 2013, that other assurance and other non-audit services are positively associated with discretionary accruals, while tax services are not significantly related to audit quality⁸⁶.

Gros (2016) runs a regression on a sample of German companies, subject to enforcement examinations over the period 2005-2012, using the natural logarithm of the fees paid for other assurance services, tax services and other non-audit services as well as the ratio of each category to total fees⁸⁷. His results show no significant association between the level or type of NAS fees and the existence of enforcement findings. Conversely, the results of *Molls* (2013), show that services from the category b) and d) are negatively associated with audit quality, while tax services are irrelevant⁸⁸. She uses a sample of German companies listed on the regulated market between

⁸² Krauß/Zülch (2013), p. 321.

⁸³ Krauß/Zülch (2013), p. 317.

⁸⁴ Krauß/Zülch (2013), p. 318.

⁸⁵ Krauß/Zülch (2013), p. 319.

⁸⁶ Hohenfels/Quick (2018), OnlineFirst 25 October 2018. Available at: <https://link.springer.com/article/10.1007/s11846-018-0306-z>.

⁸⁷ Gros (2016), pp. 220, 238.

⁸⁸ Molls (2013), p. 226.

2005 and 2010 and analyses whether each category has a different impact on audit quality, measured with discretionary accruals⁸⁹. She offers the following interpretation of her results: it is possible that since tax services are regulated by the tax consultant fee regulation (Steuerberatungsgebührenverordnung), this might leave limited room to earn quasi-rents⁹⁰. As to why other assurance services are negatively associated with audit quality, she emphasises that in most observations in her sample, valuation services are still included under category b) (later d)), which means that category b) does not include merely other assurance services⁹¹.

Finally, *Quick/Sattler* (2011b) and *Sattler* (2011) analyse the relation between NAS fees and auditor independence for a sample of German companies listed on the DAX, MDAX, SDAX and TecDax for the years 2005 to 2007. They find no significant association between the provision of other assurance services and audit quality, measured with discretionary accruals, so that they assume that knowledge spillovers and economic bonding offset each other⁹². Conversely, they find that fees for tax services are positively correlated with income-increasing discretionary accruals and that fees for other non-audit services are positively associated with discretionary accruals. These results might suggest that for these services the economic bonding argument dominates over the positive externalities of knowledge spillovers⁹³.

3.3.2.5 Client importance

Previous studies have also analysed the risk posed by client-specific quasi-rents in relation to all quasi-rents the auditor earns from other clients or in other words the economic importance of the client. Client importance is conventionally measured as the ratio of the fees paid by a client to the auditor's total revenues, and sometimes as the ratio of NAS fees to the auditor's total revenues. *Fiallo/Hecker* (2019b) find, for a sample of German companies falling under the enforcement regime, some indications that companies with a fee-to-revenue ratio higher than 15% are more likely to have enforcement findings⁹⁴. This result seems to suggest that client importance might be

⁸⁹ Molls (2013), pp. 116-117.

⁹⁰ Molls (2013), p. 216.

⁹¹ Molls (2013), p. 216. The reclassification of valuation services under category d) has followed the introduction of BilMoG, see here the considerations made in 3.4.2.

⁹² Quick/Sattler (2011b), p. 327; Sattler (2011), pp. 402-404.

⁹³ Quick/Sattler (2011b), pp. 327-336; Sattler (2011), pp. 402-417.

⁹⁴ Fiallo/Hecker (2019b), pp. 295-296. The authors run a univariate conditional logistic regression. The coefficient of the variable of interest is significant at the 10% level. As a robustness check, company's size, measured respectively as total assets and the natural logarithm of total assets, is added as a control variable. The coefficient of the main variable maintains the same sign and significance.

negatively associated with audit quality, which is why setting a limit to the fees an auditor can obtain from a client might be helpful to improve auditor independence.

Lopatta/Kaspereit/Canitz/Maas (2015) find a negative association between client importance and audit quality⁹⁵. Their results show, for a sample of German listed companies over the period 2005-2011, a positive association between the amount of income-increasing discretionary accruals and client importance, measured as the ratio of total fees to auditor's revenues and the ratio of NAS fees to auditor's revenues. They conclude that there is evidence of impaired independence for auditors of economically important clients, who engage in income-increasing earnings management.

Other authors find no association between client importance and accounting quality⁹⁶. *Quick/Sattler* (2011a), who analyse German companies listed on the DAX, MDAX, SDAX and TecDax between 2005 and 2007, argue that the lack of association is an indication that client importance does not impair independence at least below the 15% threshold prescribed by the German regulator⁹⁷. They conclude that while the 15% limit seems to prevent independence issues, it is difficult to assess if exceeding the threshold leads to lower audit quality as but few companies in the sample have a fee to revenue ratio higher than 15%⁹⁸. Further, *Sattler* (2011) observes, for the same sample of German listed companies, no significant association between the ratio of NAS fees to the auditor's revenue and discretionary accruals, concluding that client importance does not affect audit quality⁹⁹.

The study of *Zimmermann* (2008) also shows, for a sample of German companies listed in 2005 on the DAX, TecDAX, SDAX, MDAX und GEX, no significant association between client importance and discretionary accruals¹⁰⁰. However, she highlights two limitations of her study, namely: the research sample only includes companies audited by the Big 5 (BDO, Deloitte & Touche, EY, KPMG, PwC), as data for smaller auditors are not available; and due to the German

⁹⁵ Lopatta/Kaspereit/Canitz/Maas (2015), pp. 580-583.

⁹⁶ Krauß/Zülch (2013), p. 321; Zimmermann (2008), p. 220; Sattler (2011), p. 415; Quick/Sattler (2011a), pp. 84-85.

⁹⁷ The limit was regulated by sec. 319a para. 1 no. 1 HGB old version and it is now included in art. 4 sec. 3 EU-reg.

⁹⁸ Quick/Sattler (2011a), pp. 84-85.

⁹⁹ Sattler (2011), pp. 412-416.

¹⁰⁰ Zimmermann (2008), pp. 218-221.

regulation the proportion of revenues obtained by a client in the sample never exceeds 15%, leaving an information gap on the effect on audit quality above this level.

Finally, *Krauß/Zülch* (2013) find in their analysis of major listed companies for the period 2004-2011 that although NAS fees, especially for other assurance services, are positively associated with discretionary accruals, there is no significant association between client importance and discretionary accruals¹⁰¹. They argue that measuring client importance at firm level does not properly address the fee dependence issue, as partners may be more concerned about the profitability of their own engagements, rather than the firm's reputation. Similar considerations have been expressed by *Quick/Sattler* (2011a), which question if it would be more appropriate to set limitations to the revenues the auditor receives from a client at office or partner level¹⁰².

3.3.3 Research hypotheses

There are two aspects to consider when assessing the extent to which the provision of NAS affects audit quality. On the one hand, the additional knowledge gained during the provision of NAS might increase the ability of the auditor to detect errors¹⁰³, on the other hand, the provision of NAS might generate additional quasi-rents (also in the form of cost savings), which might strengthen the economic bond and lead to impaired independence¹⁰⁴. How audit quality is affected by NAS depends on which effect dominates¹⁰⁵. It is also possible that both effects offset each other. Or that the provision of NAS has no impact at all on audit quality¹⁰⁶. Even though the auditing theory and the empirical results so far have not solved this dilemma, the EU regulator assumes that the

¹⁰¹ Krauß/Zülch (2013), p. 321.

¹⁰² Quick/Sattler (2011a), p. 85.

¹⁰³ The potential benefits of knowledge spillovers on audit quality are also mentioned in Fiallo/Hecker (2019a), p. 231. A more detailed analysis of the theoretical arguments beyond knowledge spillovers is included in sec. 3.3.1.2 of this study.

¹⁰⁴ DeAngelo (1981a), p. 113. On the interaction between auditor independence (or lack thereof) and audit quality, see e.g. Fiallo/Hecker (2019a), p. 226.

¹⁰⁵ It also depends on how the two effects interact with each other. As already mentioned in Fiallo/Hecker (2019a), pp. 226 and 231: DeAngelo (1981a), p. 116, points out that the ability of the auditor to detect errors is also conditional to his willingness to report them; also, Beck/Frecka/Solomon (1988), pp. 58-59, question the assumption that NAS impair auditor independence, as they argue that under certain circumstances, recurring NAS might decrease the economic bonding.

¹⁰⁶ E.g. Francis (2004), p. 363, questions the existence of knowledge spillovers within (big) audit firms, where audit and non-audit services are likely to be provided by different teams. Similarly, one might also question the risk of economic bonding within audit firms, where the audit engagement partner might not have an economic interest in the provision of NAS, if this is taken care of by another department (and another partner). Regarding the transfer of knowledge, Francis (2004), p. 363, adds that, even among bigger audit firms, services like the annual tax declaration are often provided by the audit team as they have a direct impact on the related tax accounts, which means they might be particularly suitable for knowledge spillovers.

provision of certain NAS compromises auditor independence¹⁰⁷. The EU-regulation does not prescribe a general ban on NAS but lists those services, which cannot be provided by the statutory auditor. This decision might be interpreted with the regulator believing that, while certain services are considered risky for auditor independence, other services might be innocuous or even beneficial to audit quality.

With the enactment of the Accounting Law Reform Act (Bilanzreformgesetz) in 2005, it became mandatory for German listed companies to disclose fees paid to the auditor for audit and non-audit services¹⁰⁸. According to sec. 314 para. 1 no. 9 HGB, companies must disclose the fees charged by the group auditor according to four service categories¹⁰⁹: a) statutory audit of financial statements, b) other assurance services, c) tax services and d) other non-audit services. Category a) includes any services employed within the statutory audit. Assurance services, as defined in art. 2 para. 1 Public Accountant Act (Wirtschaftsprüferordnung - WPO), which do not fall under the statutory audit, are reported under category b). They are mainly not subject to the limitations of the EU-regulation, and if they are required by Union or national legislation, they are not subject to the cap¹¹⁰. Category c) encompasses all services related to tax advisory e.g. the annual tax declaration and tax planning. Tax services are allowed only if they have no direct or material effect on the audited financial statements, which must be documented and explained in the auditor's report¹¹¹. All services that are not classified under the categories a), b) or c) are included in category d) other non-audit services. Since the introduction of the Accounting Law Modernization Act (BilMoG), which affected financial years beginning after 31 December 2008, valuation services are included

¹⁰⁷ Recital 8, EU-reg.

¹⁰⁸ As mentioned in Fiallo/Hecker (2019a), p. 228, the legal obligation is for financial years beginning after 31 December 2004, according to art. 58 para. 3 of the Introductory Act to the German Commercial Code (Einführungsgesetz zum Handelsgesetzbuch - EGHGB).

¹⁰⁹ Simon-Heckroth/Lüdders (2017), pp. 250-252, provide a detailed account of which services fall under each category. For a comparison of the provisions included in sec. 314 para. 1 no. 9 HGB before and after the introduction of the Accounting Law Modernization Act (BilMoG), see Fiallo/Hecker (2019a), pp. 229-230.

¹¹⁰ Fiallo/Hecker (2019a), p. 232; Fiallo/Hecker (2019b), p. 292.

¹¹¹ Additionally, the auditor must comply with the principles of independence outlined in Directive 2006/43/EC. See art. 5 para. 3 EU-reg. and sec. 319a para. 1 HGB. On the Member State option and the choice of the national regulator to allow some tax services, see Fiallo/Hecker (2019a), p. 230.

in category d)¹¹². Other non-audit services are subject to wide restrictions, which have been further extended by the EU-regulation¹¹³.

Different restrictions according to the type of NAS might be motivated by the different risk these services pose to auditor independence and by the potential benefit they might have on audit quality. Other assurance services might be more closely related to the statutory audit than tax and other non-audit services, which means they might be the most suitable for generating knowledge spillovers¹¹⁴. Among tax services, knowledge spillovers might arise when preparing tax declarations, as they have a direct impact on the tax accounts in the financial statements¹¹⁵. However, the existence of knowledge spillovers does not always lead to higher audit quality, as any cost savings might increase the quasi-rents¹¹⁶.

Since other assurance services are subject to fewer restrictions and are largely diffused, the risk of economic bonding could be higher for them than for other categories¹¹⁷. For example, *Ratzinger-Sakel/Schönberger* (2015) find that the fee ratios for DAX companies in 2013 were 39.2% for other assurance services, 10.7% for tax services and 13.9% for other non-audit services, indicating that on average companies pay significantly more fees for other assurance services than for tax and other non-audit services¹¹⁸. If this is always the case, remains to be seen¹¹⁹. Also, the stricter regulation on tax and other non-audit services might prevent these services from significantly contributing to economic bonding¹²⁰. In conclusion, it is unclear whether a certain category has a

¹¹² Before BilMoG valuation services were disclosed under category b). For a description of the changes in the disclosure of auditor's fees before and after BilMoG, see Fiallo/Hecker (2019a), pp. 229-230.

¹¹³ It is mostly acknowledged that the introduction of the "black list" has increased the restrictions to the provision of NAS in Germany, see Ratzinger-Sakel/Schönberger (2015), p. 66 and Fiallo/Hecker (2019a), p. 230.

¹¹⁴ Krauß/Zülch (2013), pp. 310-311; Sattler (2011), pp. 365-366; Quick/Sattler (2011b), p. 317.

¹¹⁵ Francis (2004), p. 363.

¹¹⁶ Beck/Frecka/Solomon (1988), pp. 57-61; Ostrowski/Söder (1999), pp. 559-560.

¹¹⁷ Krauß/Zülch (2013), p. 310.

¹¹⁸ Ratzinger-Sakel/Schönberger (2015), pp. 69-70. Fee ratios are measured as the proportion of NAS fees (for each NAS category) to audit fees, in a given year.

¹¹⁹ Calculating the fee ratios for the research sample shows that in the error/matching year companies have paid most for other non-audit services (28%; control group: 19%), followed by other assurance services (22%; control group: 19%) and by tax services (9%; control group: 13%). Also, Fiallo/Hecker (2019a), p. 232, have shown that in the error year 55.2% of the companies in the error group (control group: 71.3%) have bought other non-audit services, which is more than the proportion of companies buying other assurance services (44.8%; control group: 44.1%) and tax services (41.3%; control group: 63.6%).

¹²⁰ Krauß/Zülch (2013), pp. 310-311.

positive or a negative impact on audit quality, which is why the following hypothesis is formulated in the null form.

H1: The provision of other assurance services/tax services/other non-audit services is not associated with audit quality.

Besides the concerns related to the type of NAS, the preamble to the EU-regulation argues that the level of fees and their structure might impair auditor independence¹²¹. To prevent NAS fees becoming too high, the EU-regulation has introduced a cap. According to art. 4 para. 2 EU-reg., the fees for NAS in the current period cannot exceed 70% of the average audit fees charged by the auditor in the last three consecutive years¹²². The cap applies only to PIEs, which have purchased NAS in each of the previous three years from the incumbent auditor. It follows that, if the company switches auditor, or does not qualify as a PIE in every year or does not purchase NAS in each of the previous three years, then the cap does not apply¹²³.

In general, the introduction of a cap is reasonable, when there is a non-linear relation between the level of NAS fees and audit quality, so that auditor independence becomes impaired only after exceeding a certain threshold¹²⁴. In the case of the EU-regulation, 70% is the cut-off value above which NAS fees are believed to decrease audit quality, while NAS fees below the threshold should be uncritical. Due to the lack of theoretical arguments as well as of specific empirical evidence supporting the 70% threshold, this choice seems arbitrary and questions the ability of the cap to improve audit quality¹²⁵.

Prior studies have tried to analyse whether the regulator should limit NAS by setting a cap on the proportion of NAS fees. *Quick/Warming-Rasmussen* (2009) report that German investors perceive a ratio of NAS fees to total fees of 25% as critical¹²⁶. In response to this, *Quick/Sattler* (2011b) analyse, for a sample of German companies listed on the DAX, MDAX, SDAX and TecDax for

¹²¹ Recital 7 EU-reg.

¹²² For a more detailed analysis of the provisions included in art. 4 para. 2 EU-reg., see Fiallo/Hecker (2019b), pp. 290-291.

¹²³ These criteria lead to a significant decrease in the number of companies to which the cap applies. See e.g. Fiallo/Hecker (2019b), p. 292.

¹²⁴ Quick/Sattler (2011b), p. 333.

¹²⁵ Ratzinger-Sakel/Schönberger (2015), pp. 62-63; Pott/Schröder/Weckelmann (2014), p. 502. These doubts are also briefly discussed in Fiallo/Hecker (2019b), p. 291.

¹²⁶ Quick/Warming-Rasmussen (2009), p. 142.

the years 2005 to 2007, if exceeding the 25% threshold, as well as a more conservative 10% threshold, leads to impaired independence of mind¹²⁷. Their results show that a proportion of NAS to total fees larger than 25%, but also larger than 10%, is positively associated with the amount of absolute discretionary accruals, which suggests that NAS fees might be a threat to independence already at the 10% value.

Krauß/Zülch (2013) replicate the described analysis on a sample of major listed companies for the period 2004-2011¹²⁸. They find that companies exceeding the 25% threshold are more likely to have larger absolute and positive discretionary accruals. Regarding the 10% threshold, they find no significant association with audit quality, which might indicate that auditor independence is impaired only at higher levels of NAS.

More recently, *Hohenfels/Quick* (2018) find a negative relation between NAS fees and audit quality, measured with discretionary accruals, even when NAS fees are below the 70% cap, which they interpret as the cap being too high¹²⁹. Additionally, they estimate, for a sample of German listed companies on the CDAX between 2006 and 2013, an optimal cap on NAS fees of 41%. In contrast to the present study, however, auditor changes and the provision of NAS in each year during the three-year period do not appear to have been considered¹³⁰. The same seems to apply to *Holm* (2016), who analyses the relationship between audit and non-audit fees for companies from five European countries with an NAS fee to audit fee ratio below and above 70%¹³¹. He postulates that a positive association between audit and non-audit fees shows the existence of economies of scope. If the auditor retains at least some of the cost savings generated by the joint provision of services, this could increase economic bonding and impair his independence. The author collects data for 903 German non-financial companies prior to the enactment of the EU-regulation (2010-2013). His results show that in Germany the relation between audit and non-audit fees is positive below and above 70%, indicating that economic bonding might exist also below this threshold and that limiting NAS fees might have the potential to reduce the independence problem. However, when measuring the cap as the ratio of NAS to the average audit fees charged in the prior three

¹²⁷ Quick/Sattler (2011b), pp. 333-334.

¹²⁸ Krauß/Zülch (2013), p. 320.

¹²⁹ Hohenfels/Quick (2018), OnlineFirst 25 October 2018. Available at: <https://link.springer.com/article/10.1007/s11846-018-0306-z>.

¹³⁰ Similar considerations are made by Fiallo/Hecker (2019c), p. 348 footnote no. 161.

¹³¹ Holm (2016), pp. 2-6.

years, the results are not significant, thus failing to provide any evidence on whether the cap improves audit quality¹³².

Köhler/Theis (2018) estimate the impact of the cap by calculating for the years 2005 to 2015 the ratio of NAS fees to audit fees for companies with securities admitted to trading on a regulated market¹³³. They find that, while DAX companies have since 2011 an average ratio higher than 70%, smaller companies remain well under 60%, which indicates a potentially limited impact of the cap. Finally, *Ratzinger-Sakel/Schönberger* (2015) calculate for DAX companies the ratio of NAS fees for 2013 to the average audit fees from prior periods (2010-2012) and find it to be above 70% in 33% of cases¹³⁴. They conclude that most companies in their sample already pay NAS fees below 70%, which questions the ability of the cap to have a significant impact on the audited companies.

In this study, I identify which companies in the sample would have violated the cap, if the EU-regulation had already applied. Given the lack of empirical evidence and especially the lack of theoretical arguments justifying the 70% threshold, the following hypothesis is formulated in the null form.

H2: The violation of the 70% cap as prescribed by art. 4 para. 2 EU-reg. is not associated with audit quality.

Another concern of the EU regulator is that of the auditor becoming excessively dependent on a client¹³⁵. Therefore, art. 4 para. 3 EU-reg. aims at preventing an auditor receiving more than 15% of his revenue from one entity for a longer period¹³⁶. The auditing theory argues that the existence of quasi-rents coming from different clients is a “collateral against opportunistic behavior”, as the auditor will not risk damaging his reputation for fear of losing other contracts¹³⁷. However, if the economic interest related to a client exceeds the expected loss in quasi-rents from other clients, the auditor will compromise his independence to retain the major client. It follows that the auditor’s

¹³² Holm (2016), p. 41 endnote no. 7.

¹³³ Köhler/Theis (2018), pp. 407-408.

¹³⁴ Ratzinger-Sakel/Schönberger (2015), p. 69.

¹³⁵ Recital 7 EU-reg.

¹³⁶ For a more detailed analysis of the provisions included in art. 4 para. 3 EU-reg., see Fiallo/Hecker (2019b), p. 294.

¹³⁷ DeAngelo (1981b), p. 197.

behaviour is affected by the economic importance of the client or the ratio of the client-specific quasi-rents to the quasi-rents from all clients¹³⁸.

Assuming a positive correlation between quasi-rents and fees, the economic importance can be approximated by the ratio of total fees divided by the auditor's revenue¹³⁹. As in the case of the cap, the decision to set a threshold suggests that the relation between client importance and audit quality is of a nonlinear nature, with audit quality being threatened by the excessive economic dependence on a client only after exceeding a certain threshold¹⁴⁰. Based on the theoretical arguments and focusing on the critical value set by the European regulator, I estimate that a client is important when he contributes to more than 15% of an auditor's revenue and formulate the following hypothesis.

H3: Client importance is negatively associated with audit quality.

3.4 Research design

3.4.1 Surrogates for audit quality

According to *DeAngelo* (1981b), audit quality is the auditor's perceived ability to detect misstatements and to report them¹⁴¹. While the ability to detect misstatements depends on the auditor's competence and level of effort, the decision to report them is conditional on his independence. Since both the procedures performed by the auditor and the specific client-auditor dynamic cannot be observed from the outside, measuring audit quality has proven to be a challenge¹⁴². In an attempt to develop a framework for studying audit quality, *Francis* (2004) proposes to visualise audit quality as a continuum, where the lower and upper extremities imply a very low or a very high level of audit quality¹⁴³. He then describes the lowest point of this continuum as characterised by the presence of audit failures, which occur when the auditor does not comply with the applicable standards and fails to issue the appropriate audit opinion¹⁴⁴. While

¹³⁸ DeAngelo (1981b), pp. 187-192.

¹³⁹ E.g. Ostrowski/Söder (1999), p. 558; Chung/Kallapur (2003), p. 932.

¹⁴⁰ Recital 7 EU-reg.

¹⁴¹ DeAngelo (1981b), p. 186. This definition provided by DeAngelo is commonly used in auditing literature, see e.g. Fiallo/Hecker (2019a), p. 226.

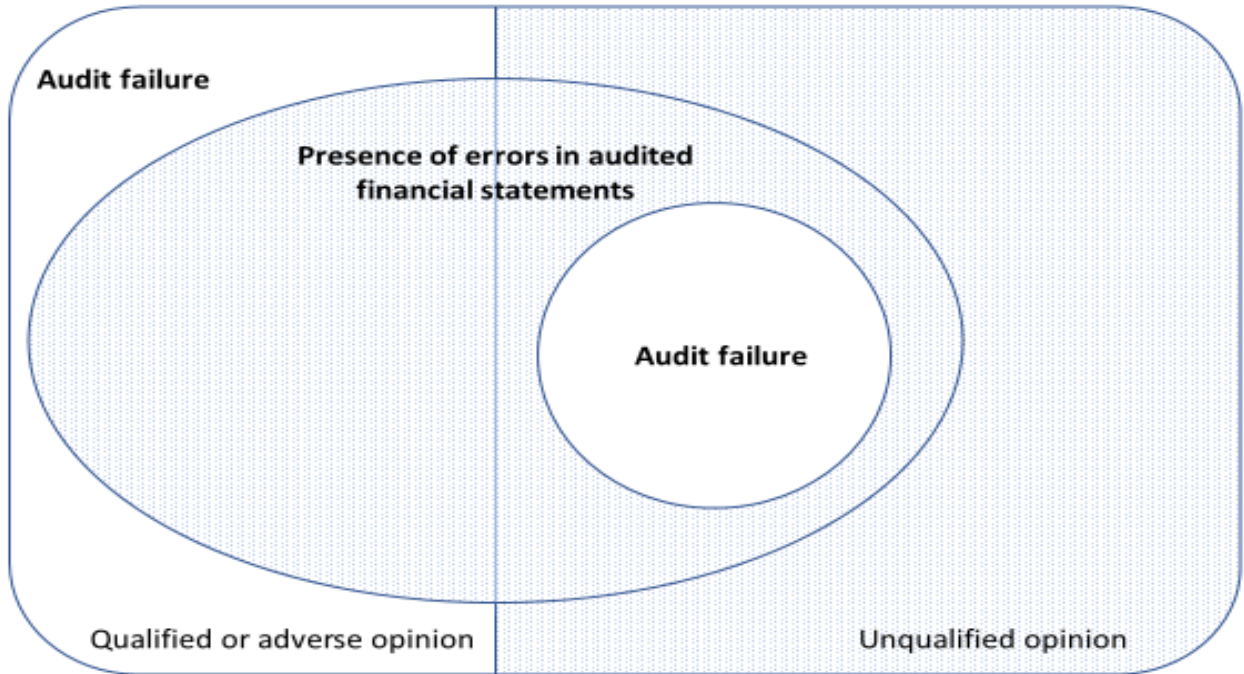
¹⁴² DeAngelo (1981b), p. 186. Similar considerations are made e.g. by Fiallo/Hecker (2019a), p. 226.

¹⁴³ Francis (2004), p. 346; Francis (2011), p.129.

¹⁴⁴ Francis (2004), p. 346; Francis (2011), p. 127.

the definition provided by *DeAngelo* (1981b) appears abstract, the definition of *Francis* (2004) allows at least for a concretisation of the concept of low audit quality¹⁴⁵.

Figure 3.1: Definition of audit failure



As shown in Figure 3.1, an audit failure takes place when an auditor issues a qualified or adverse opinion even though the financial statements are free from material errors (false positive), and most notably when he issues an unmodified opinion in the presence of material errors (false negative). In this context, the presence of accounting errors in audited financial statements and the lack of a qualified or adverse opinion might be interpreted as an audit failure and be used as an indicator of low audit quality¹⁴⁶. There is, however, a caveat. As shown in Figure 3.1, there is an area where, despite the presence of errors and of an unmodified opinion, we cannot speak of audit failures. Due to the inherent limitations of the audit procedures and the residual audit risk, the auditor can only attest with reasonable assurance that financial statements are free from material errors¹⁴⁷. As argued by *Knechel* (2009), audit risk can be reduced to an “appropriate” level, but perfect assurance cannot be attained, which is why errors may go undetected¹⁴⁸. It follows that the presence of accounting

¹⁴⁵ Knechel/Krishnan/Pevzner/Schefchik/Velury (2013), pp. 387-388; Knechel (2009), pp. 5-6.

¹⁴⁶ Tritschler (2013), p. 4.

¹⁴⁷ ISA 240.5. See here also the discussion in Fiallo/Hecker (2019a), p. 227.

¹⁴⁸ Knechel (2009), pp. 6-7.

errors in financial statements where the auditor issued an unmodified opinion does not univocally define an audit failure. However, an audit failure (false negative) requires the presence of errors and of an unmodified opinion, which means that enforcement findings and audit quality should be strongly negatively correlated¹⁴⁹.

Notwithstanding their limitations, enforcement findings might be a better surrogate for audit quality than other popular proxies such as auditor's modified/going concern opinions or discretionary accruals¹⁵⁰. *Sharma/Sidhu* (2001) argue that interpreting modified opinions as signals for independence and unmodified opinions as a lack thereof is misleading, unless it was possible to determine which opinion would have been appropriate¹⁵¹. Due to the evident difficulties for researchers to determine the appropriate opinion, some studies prefer using, as a surrogate for audit quality, the probability of issuing a GCO to companies in financial distress, or which have later filed for bankruptcy¹⁵². However, GCOs also have limitations. The auditor's decision to issue a GCO relies on estimates and predictions, which might be undermined by the occurrence of unforeseeable events (e.g. global crisis)¹⁵³. It follows that failing to issue a GCO to a company in financial difficulties does not necessarily imply a lack of independence. Especially when, as argued by *Ewert* (2004), the auditor does not expect any future economic benefit from the client (e.g. in case of bankruptcy)¹⁵⁴.

Another popular proxy for audit quality is discretionary accruals. Accruals are the difference between earnings and operating cash-flow¹⁵⁵. Accruals might arise from normal business activities. However, they might also reflect earnings management conducted by managers. The accounting

¹⁴⁹ It should be noted that the presence of errors in audited financial statements is not equal to the existence of enforcement findings. Rather, financial statements subject to an enforcement announcement are a subset of the population of financial statements with errors. As to why this occurs, see the discussion in *Fiallo/Hecker* (2019a), pp. 226-227. Also, *Fiallo/Hecker* (2019a) argue that this consideration does not contradict the use of enforcement findings to proxy for low audit quality, but it rather indicates that the lack of enforcement findings is not equal to the lack of errors.

¹⁵⁰ A recent review of auditing research has shown that modified or going concern opinions and discretionary accruals, are the most used proxies for audit quality. See *Simnett/Carson/Vanstraelen* (2016), p. 14. Examples of studies using going concern opinions are *Blay/Geiger* (2013) and *Ratzinger-Sakel*, (2013). Examples of studies using discretionary accruals are *Lopatta/Kaspereit/Canitz/Maas* (2015), *Krauß/Zülch* (2013), *Krauß/Pronobis/Zülch* (2015), *Molls* (2013), *Quick/Sattler* (2011b), *Sattler* (2011) and *Tebben* (2011). For an extensive list of proxies for audit quality, see e.g. *Gros/Worret* (2014).

¹⁵¹ *Sharma/Sidhu* (2001), p. 597.

¹⁵² E.g. *Ratzinger-Sakel* (2013), p. 130.

¹⁵³ *Tritschler* (2013), p. 66.

¹⁵⁴ *Ewert* (2004), p. 257. A similar consideration is made by *Blay/Geiger* (2013), p. 583.

¹⁵⁵ E.g. *Wagenhofer/Ewert* (2015), p. 280.

literature has therefore produced different models, which try to discriminate between normal accruals, which are compatible with the company's performance, industry, and other economic characteristics, and discretionary accruals, which result from the arbitrary decisions of managers¹⁵⁶. Since one of the auditor's tasks is to restrict earnings management practices, discretionary accruals can be indirectly used to measure audit quality¹⁵⁷. This approach, although widely used in prior auditing studies¹⁵⁸, is often criticised, as it has a few drawbacks. First, earnings management is essentially a measure of accounting quality and only an indirect measure of audit quality¹⁵⁹. Second, earnings management does not always imply the presence of accounting errors. In fact, earnings management can be practised within the scope of GAAP, in which case the auditor would have no reason to issue a modified opinion¹⁶⁰. Third, the level of "normal" accruals is usually estimated using a cross-sectional approach, where companies are clustered by sector and year. *Jackson (2018)* states that "econometrically speaking, discretionary accruals are simply deviations from industry averages", while attributing this difference entirely to earnings management is an overconfident assumption¹⁶¹. In fact, discretionary accruals are very sensitive to the estimation sample and to the characteristics of peer firms (e.g. the presence of earnings management in the estimation sample), which might increase the risk of incorrectly classifying accruals, thus leading to measurement errors¹⁶².

In conclusion, each surrogate has its limitations¹⁶³. In this paper, I choose to use the existence of enforcement findings as a proxy for audit quality, as the presence of errors in audited financial statement should be a good indicator of low audit quality¹⁶⁴. Additionally, the fact that enforcement findings are established by an independent institution, it allows to minimise measurement errors¹⁶⁵.

¹⁵⁶ Ronen/Yaari (2008), p. 372.

¹⁵⁷ Gros/Worret (2014), p. 348.

¹⁵⁸ E.g. Quick/Sattler (2011b), p. 320; Krauß/Zülch (2013), p. 311; Sattler (2011), p. 210.

¹⁵⁹ Quick/Schmidt/Simons (2016), p. 197.

¹⁶⁰ Quick/Schmidt/Simons (2016), p. 196; Gros/Worret (2014), p. 348.

¹⁶¹ Jackson (2018), pp. 137-138.

¹⁶² Jackson (2018), pp. 140-144, conducts an experiment where he intentionally "adds" earnings management to one of the companies in the estimation sample. He is able to show that, since discretionary accruals are estimated within the group, the modification in the fundamentals of one company leads to different values of discretionary accruals for all the companies in the group.

¹⁶³ For an extensive analysis, see also DeFond/Zhang (2014), pp. 283-289.

¹⁶⁴ A similar assumption has been made by e.g. Frey/Möller/Weinzierl (2016), p. 563 and Gros (2016), pp. 219-220.

¹⁶⁵ Hoehn/Strohmer (2013), p. 10; DeFond/Zhang (2014), p. 284; Dechow/Ge/Schrand (2010), p. 371. Dechow/Ge/Schrand (2010), p. 371, caution that although using enforcement releases leads to less type I errors (classifying financial statements as erroneous when they are not), it might introduce a selection bias. The bias arises

3.4.2 Sample selection

When financial statements are found to be erroneous during an enforcement examination and this information is relevant to the financial market, the company is required to publish the enforcement findings in the Federal Gazette, as well as in a multi-regional financial newspaper or an additional electronic financial information database¹⁶⁶. To build the research sample, financial statements subject to enforcement findings are identified based on the enforcement announcements published in the Federal Gazette (Bundesanzeiger)¹⁶⁷. The search from July 2005, which marked the beginning of the enforcement examinations, until 31 December 2017, yielded 258 erroneous financial statements¹⁶⁸.

As shown in Table 3.1, double announcements or amendments of already released announcements are excluded from the sample, as they refer to the same set of financial statements. Further, announcements regarding half-year reports are eliminated as they are not subject to a mandatory audit¹⁶⁹. Foreign companies are also excluded, as they may apply different provisions to those prescribed by German GAAP regarding the disclosure of auditor's fees. In the interest of comparability, errors found in annual financial statements and in consolidated financial statements not reporting under IFRS are eliminated from the sample. Since I am interested in analysing whether there is a link between the provision of NAS and low audit quality, I eliminate observations regarding financial statements for which the auditor issued a qualified or adverse opinion¹⁷⁰. Finally, as audit and non-audit fees are disclosed for the first time in financial statements beginning

when the choice of the companies to be examined is not completely random like in the case of the FREP, where e.g. companies listed on an index are examined every 4 to 5 years, while other companies every 8 to 10 years. For the description of the sampling approach of the FREP, see section 5.2.2.

¹⁶⁶ Sec. 109 para. 2 of the German Securities Trading Act (WpHG). See here also e.g. Hecker/Wild (2012), pp. 6-7 as well section 5.2.

¹⁶⁷ The same sample approach has been used in Fiallo/Hecker (2019a), p. 228. Compared to the sample described in Fiallo/Hecker (2019a), p. 228 Table 2, there are 10 fewer observations. In this study, 7 announcements regarding financial years beginning before 1 January 2005 are eliminated, as companies were not required to disclose auditor fees. Also, 3 announcements are eliminated as fee data were not available. These are one set of financial statements for the year 2005, where the fees charged to the group auditor were not disclosed and two sets of financial statements for the years 2005 and 2015, which only reported the fees paid to the worldwide network of the group auditor.

¹⁶⁸ The search on the website of the Federal Gazette (<https://www.bundesanzeiger.de>) for the chosen period produces 250 results. As explained in the note to Table 2 in Fiallo/Hecker (2019a), p. 228, eight enforcement announcements regarding different sets of financial statements (e.g. yearly and half-yearly) are here separately accounted for.

¹⁶⁹ Fiallo/Hecker (2019a), p. 226.

¹⁷⁰ As to why this is done, see sec. 3.4.1 as well as the discussion in Fiallo/Hecker (2019a), p. 226.

after 31 December 2004¹⁷¹, I exclude those beginning before this date and additionally three observations with missing fee data¹⁷². A final sample of 141 erroneous financial statements for financial years between 2005 and 2015 remains.

Table 3.1: Sample selection

Enforcement announcements issued in the Federal Gazette until 31/12/2017	258
Announcements disclosed twice or amended	- 17
Announcements regarding half-year financial statements	- 30
Announcements regarding foreign companies	- 19
Announcements regarding annual financial statements (local GAAP)	- 11
Announcements regarding consolidated financial statements (US-GAAP or local GAAP)	- 8
Announcements regarding financial statements which received an adverse auditor's opinion	- 3
Announcements regarding financial statements which received a qualified auditor's opinion	- 19
Announcements regarding financial statements beginning before 01/01/2005	- 7
Announcements regarding financial statements for which fee data were not available	- 3
Final sample	141

Notes: This table is based on Fiallo/Hecker (2019a), p. 228 Table 2. The differences in the sample size are explained in footnote 167.

In order to analyse the relationship between auditor's fees and audit quality, a control group characterised by higher audit quality must be built¹⁷³. The control sample is drawn from the population of companies with securities listed in Germany¹⁷⁴. Each error company is matched in the error year to a control company according to the following criteria¹⁷⁵: the matched company must (1) be a German company, (2) have the same industry classification, (3) prepare consolidated

¹⁷¹ As mentioned in Fiallo/Hecker (2019a), p. 228, the legal obligation is for financial years beginning after 31 December 2004, according to art. 58 para. 3 of the Introductory Act to the German Commercial Code (Einführungsgesetz zum Handelsgesetzbuch - EGHGB).

¹⁷² One set of financial statements for the year 2005 are eliminated as the company failed to disclose the fees charged to the group auditor. Further, two sets of financial statements for the years 2005 and 2015 report only the fees paid to the worldwide network of the group auditor. They are excluded for the sake of comparability.

¹⁷³ The approach employed to build the control group is the same as in Fiallo/Hecker (2019a), p. 228.

¹⁷⁴ The population includes 921 German companies, with securities traded on the national market as of December 2015, which are available in Thomson Reuters Eikon.

¹⁷⁵ For details on the matching approach, see Fiallo/Hecker (2019a), p. 228. In footnote no. 32, Fiallo/Hecker (2019a) discuss the trade-off between matching accuracy and correlation issues. The sample includes non-independent observations, due to the fact that the chosen matching approach allows a company to be sampled more than once in different years (partial replacement). Prior studies suggest, in the case of non-independent observations, to weight the repeated observations and to control for the correlation among standard errors (e.g. Shipman/Swansquist/Whited (2017), p. 217; Stuart (2010), p. 9; Armstrong/Jagolinzer/Larcker (2010), p. 244). However, this does not seem necessary, when employing a conditional logistic regression clustered at pair-level, as any correlation within a cluster is already accounted for. As for the correlation among clusters, there are three pairs, which are each repeated twice in the sample. In order to make sure that the correlation among clusters does not affect the results, the analyses are re-run without the six firm/year observations. The results do not significantly change.

financial statements according to IFRS¹⁷⁶, (4) be subject to the enforcement system in the matching year and in the following year (5) not have any enforcement announcements as of 31 December 2017, (6) be similar in size, measured as the minimum absolute difference in total assets. By falling under the enforcement regime and not having had any enforcement findings so far, control companies are expected to have a higher audit and accounting quality compared to error companies. Nonetheless, it cannot be stated with certainty that companies in the control group are free from accounting errors¹⁷⁷.

Table 3.2: Sector and year distribution

Sector	No.	Perc.	Year	No.	Perc.
Consumer Discretionary	62	22%	2005	34	12%
Financials	68	24%	2006	50	18%
Telecom Services	8	3%	2007	24	9%
Materials	10	4%	2008	38	13%
Information Technology	56	20%	2009	40	14%
Consumer Staples	8	3%	2010	16	6%
Health Care	20	7%	2011	14	5%
Industrials	44	16%	2012	30	11%
Utilities	6	2%	2013	10	4%
			2014	14	5%
			2015	12	4%
Total	282	100%		282	100%

Notes: Companies are classified according to the Global Industry Classification Standard (GICS).

Table 3.2 shows the sector distribution in the sample. Almost one fourth of the companies (68 obs.) are in the financial sector (credit institutions, insurance and real estate companies). Prior accounting studies tend to exclude financial companies from the analysis, mainly because certain accounting values are not comparable (e.g. total assets, revenue). Also, due to a more extensive regulation, fees charged by the auditor of financial companies could be higher than fees charged by auditors operating in other sectors¹⁷⁸. By conducting a logistic regression conditional on matched-pairs, observations are compared within the respective industries, which means that sector differences should not be an issue. However, to make sure that the inclusion of financial companies does not

¹⁷⁶ In one case, both annual and consolidated financial statements for the year 2005 were not available. In this case, the comparative figures according to IFRS as well as the auditor's fees corresponded in 2005 are retrieved from the 2006 consolidated financial statements.

¹⁷⁷ This issue is discussed in Fiallo/Hecker (2019a), pp. 226-227.

¹⁷⁸ E.g. Lehrbass/Scheipers (2017), p. 1439; Gros (2016), p. 230.

drive the results, sensitivity tests are performed in section 3.5.3, where all financial companies are dropped from the sample.

Table 3.2 also shows the year distribution in the sample. In this paper, I employ fee data disclosed both before and after the introduction of BilMoG, which has affected financial years beginning after 31 December 2008¹⁷⁹. With BilMoG, the interpretation of the national disclosure requirements regarding auditor's fees, included in sec. 314 para. 1 no. 9 HGB, has changed¹⁸⁰. This heterogeneity of the data might lead to measurement errors, which in turn might yield biased results. This problem is partially addressed by running an additional analysis accounting for the pre- and post- BilMoG era¹⁸¹.

3.4.3 Model and main variables

To test H1, a conditional logistic regression is employed. The conditional logistic regression is recommended for matched samples, as it allows to determine the likelihood of an outcome when the observations are not independent but paired¹⁸². The model used for H1, which has no intercept¹⁸³, is the following¹⁸⁴:

$$\begin{aligned} Error_{i,t} = & \beta_1 * NASFees_{i,t} + \beta_2 * AuditFees_{i,t} + \beta_3 * RoA_{i,t} + \beta_4 * Loss_{i,t} + \beta_5 * Lev_{i,t} \\ & + \beta_6 * CFO_{i,t} + \beta_7 * Growth_{i,t} + \beta_8 * Big4_{i,t} + \beta_9 * Change_{i,t} \\ & + \beta_{10} * IFRSfirst_{i,t} + \beta_{11} * Size_{i,t} + \varepsilon_{i,t} \end{aligned}$$

$Error^{185}$ is the binary dependent variable, which assumes the value of 1 if the financial statements for the year t of company i are subject to an enforcement announcement, and 0 otherwise. Due to

¹⁷⁹ Art. 66 para. 2 sentence 1 EGHGB.

¹⁸⁰ The changes relate to the definition of group auditor, the definition of audited company, the reference period and the categories' labels. For a comparison of the disclosure requirements regarding auditor's fees before and after BilMoG, see Fiallo/Hecker (2019a), pp. 229-230.

¹⁸¹ Similarly, in Fiallo/Hecker (2019a), p. 231 and Fiallo/Hecker (2019b), pp. 295-296, data are additionally analysed before and after the introduction of BilMoG.

¹⁸² Hosmer/Lemeshow/Sturdivant (2013), pp. 243-251; Cram/Karan/Stuart (2009), p. 480; Stuart/Shin/Cram/Karan (2013), pp. 89-90. See also the similar considerations reported by Fiallo/Hecker (2019a), p. 233.

¹⁸³ For an explanation as to why there is no intercept in the conditional logistic model, see Allison (2012), p. 195.

¹⁸⁴ In choosing the model variables, I keep in mind the rule of thumb that, to achieve a reliable estimation of the coefficients in a logistic regression, the number of covariates should stay at least in a 1:10 proportion to the least frequent outcome, which in this case is the presence (or lack) of enforcement findings. See Hosmer/Lemeshow/Sturdivant (2013), pp. 407-408. However, due to data constraints in some of the analyses the proportion is slightly under the above value.

¹⁸⁵ Hereafter, apart from Table 3.5, the variables are reported without subscripts.

the design of the control group, the value 0 indicates that the company did not have any enforcement findings in year t and in all the years subject to analysis.

The term *NAS Fees* represents the amount of NAS fees paid to the auditor and should approximate the economic interest the auditor has in the client as well as the extent of the provided services. To test if the provision of certain services affects audit quality, fees are classified according to four categories, namely fees for: a) statutory audit of financial statements, b) other assurance services, c) tax services and d) other non-audit services. Two sets of variables are built: one set is the natural log of NAS fees from categories b), c) and d) (*NAScat_b*, *NAScat_c* and *NAScat_d*)¹⁸⁶, the other set is the ratio of the fees disclosed in each category to the total fees (*NASratio_b*, *NASratio_c* and *NASratio_d*)¹⁸⁷.

I make two assumptions here. First that the fees charged by the auditor and disclosed in consolidated financial statements are a good proxy for the quasi-rents the auditor expects from the engagement¹⁸⁸. Quasi-rents are not directly observable, as besides the level of fees, they also depend on the client specific costs and their development in the future¹⁸⁹. However, I expect fees and quasi-rents to be positively correlated¹⁹⁰. Second, the fees charged by the auditor should also reflect his level of effort¹⁹¹. In this case, I expect the higher the fees, the more extensive the services provided by the auditor are.

Four different models are examined: one with all main variables (M1) and three where each NAS category is considered separately (M2, M3, M4)¹⁹². In each model, the variable *AuditFees*, which is the natural logarithm of the fees charged for the statutory audit, is included. The level of audit fees might reflect the auditor's effort to reduce audit risk to an acceptable level¹⁹³. At the same

¹⁸⁶ Previous studies using the natural log of NAS fees include e.g. Markelevich/Rosner (2013), Molls (2013), Tebben (2011) and Ratzinger-Sakel (2013).

¹⁸⁷ Previous studies for Germany that have used the ratio of NAS fees as an independent variable are e.g. Gros (2016), Krauß/Zülch (2013), Quick/Sattler (2011b), Lopatta/Kaspereit/Canitz/Maas (2015), Molls (2013), Sattler (2011) and Tebben (2011).

¹⁸⁸ Quick/Sattler (2011a), pp. 83-85.

¹⁸⁹ Wagenhofer/Ewert (2015), p. 525.

¹⁹⁰ A similar assumption has been mentioned by Wagenhofer/Ewert (2015), p. 525.

¹⁹¹ The potential existence of a fee premium is disregarded.

¹⁹² Previous studies have regressed the category variables separately (e.g. Krauß/Zülch (2013); Quick/Sattler (2011b); Sattler (2011); Gros (2016)) as well as all together (e.g. Molls (2013)).

¹⁹³ See also Simunic (1980), p. 166, who develops a model for the pricing of audit services, where he shows that, in a competitive audit market, audit fees are a function of audit risk and auditor's effort.

time, higher audit fees might increase the risk of economic bonding due to the existence of quasi-rents¹⁹⁴. Due to these opposing effects, I do not predict the sign of the coefficient of *AuditFees*.

As explained above, the national disclosure requirements regarding auditor's fees, were modified by the introduction of BilMoG, which has affected financial years beginning after 31 December 2008¹⁹⁵. One major change concerns the classification of valuation services under category d), while prior to BilMoG, they were disclosed under category b). As the association between the fees paid for services of the category b) or d) and the presence of enforcement findings might vary before and after the introduction of BilMoG, I rerun the analysis including an interaction term in the model¹⁹⁶. The dichotomous variable *Bilmog* assumes the value of 1 for financial years beginning after 31 December 2008, and 0 otherwise. The models M2 and M4 are then rerun by adding the variable *Bilmog* and an interaction term (*Interaction*) between *Bilmog* and the variables for category b) and d) (*NAScat_b*, *NAScat_d*, *NASratio_b*, *NASratio_d*)¹⁹⁷.

To test H2, I identify companies, which would have violated the provisions of art. 4 para. 2 EU-reg., if this had already applied. The variable *Cap* is a dichotomous variable assuming the value of 1, if the company is a PIE¹⁹⁸, which acquires NAS in each of the prior three years from the incumbent auditor, and which has in year t a ratio of NAS fees to the average audit fees charged in years t-1, t-2 and t-3 greater than 70%.

¹⁹⁴ DeAngelo (1981a), p. 113.

¹⁹⁵ For a comparison of the disclosure requirements before and after BilMoG, see Fiallo/Hecker (2019a), pp. 229-230.

¹⁹⁶ On the use and interpretation of interaction terms, see Hosmer/Lemeshow/Sturdivant (2013), pp. 69-70.

¹⁹⁷ On the importance of adding all constitutive terms of the interaction in the model, see Brambor/Clark/Golder (2006), pp. 66-70, 77.

¹⁹⁸ EC (2014), p. 3; IDW (2018), pp. 44-45. On the process of how it is made sure that the companies in the error and control group are PIEs in all relevant years, see Fiallo/Hecker (2019b), p. 293 footnote no. 108. As reported there, there are two exceptions, one in the error group and one in the control group. In both cases the fees paid are below the cap.

Table 3.3: Sample selection - H2

Reduced sample for the test of H2	Error group	Control group
Research sample	141	141
Observations with financial year beginning before 01/01/2008	- 56	- 56
Observations with missing fee data	- 3	- 8
Observations for which no match was available	- 10	- 5
Total	72	72
Thereof with $Cap = 1$	8 (11%)	14 (19%)
Thereof with $Cap = 0$	64 (89%)	58 (81%)

Notes: This table is partially based on Fiallo/Hecker (2019b), p. 292 Table 5. The differences in number of excluded observations are due to the different initial sample.

To verify whether auditor's fees exceed the cap in a given year, data are collected for the three prior periods. This leads to the exclusion of some observations due to lack of data¹⁹⁹. Since auditor's fees are disclosed for the first time in financial years beginning after 31 December 2004²⁰⁰, enforcement announcements regarding financial statements beginning before 1 January 2008 are eliminated as data for the three prior years are not available. Table 3.3 shows the composition of the sample employed for testing H2, which is made of 72 matched pairs. As also shown in Table 3.3, there are relatively few companies, 11% in the error group and 19% in the control group, whose fees exceed the cap in the examined years. One limitation of this result²⁰¹ is that the fees collected from financial statements and used to measure the variable Cap differ in a few aspects from the requirements of art. 4 para. 2 EU-reg.²⁰². Thus, the effect of the cap can only be approximated.

The regression model employed for testing H2 is the following:

$$\begin{aligned}
 Error_{i,t} = & \beta_1 * Cap_{i,t} + \beta_2 * RoA_{i,t} + \beta_3 * Loss_{i,t} + \beta_4 * Lev_{i,t} \\
 & + \beta_5 * CFO_{i,t} + \beta_6 * Growth_{i,t} + \beta_7 * Big4_{i,t} + \beta_8 * Change_{i,t} \\
 & + \beta_9 * IFRSfirst_{i,t} + \beta_{10} * Size_{i,t} + \varepsilon_{i,t}
 \end{aligned}$$

¹⁹⁹ The main sample, as described in section 3.4.2, is here adjusted following the same approach described in Fiallo/Hecker (2019b), p. 292 footnote no. 101 and no. 102. The final sample is the same used in Fiallo/Hecker (2019b), pp. 292-293.

²⁰⁰ Art. 58 para. 3 sentence 1 of the Introductory Act to the German Commercial Code (EGHGB).

²⁰¹ As reported in Fiallo/Hecker (2019b), p. 293 footnote no. 105, the results suffer also from the fact that some observations must be dropped from the sample as data for the matched company are not available. In 5 out of 10 observations for the error group, NAS fees exceed the cap. Conversely, in each of the 5 observations excluded from the control group fees are below the cap.

²⁰² For a detailed analysis, see Fiallo/Hecker (2019b), pp. 291-292.

To test H3, I employ the variable *ClientImp*, which is a dichotomous variable coded one, when in year *t* the ratio of fees paid by one client to the auditor's revenue exceeds 15%, and zero otherwise²⁰³. Additional data are manually collected from annual transparency reports, where auditors of listed companies (now auditors of PIEs) are required to disclose their revenue²⁰⁴. Since these data are available at the earliest for financial statements beginning in 2006²⁰⁵, several observations must be dropped from the sample. After dropping further observations for which revenue data are not found or for which no match is available, it remains a sample of 113 matched-pairs.

Table 3.4: Sample selection - H3

Reduced sample for the test of H3	Error group	Control group
Research sample	141	141
Observations with financial years for which transparency reports were not available	- 19	- 19
Observations lacking revenue data for the auditor	- 6	- 4
Observations for which no match was available ²⁰⁶	- 3	- 5
Total	113	113
thereof with <i>ClientImp</i> = 1	11 (9.7%)	4 (3.5%)
thereof with <i>ClientImp</i> = 0	102 (90.3%)	109 (96.5%)

Notes: This table is partially based on Fiallo/Hecker (2019b), p. 295 Table 7. The differences in number of excluded observations are due to the different initial sample.

Table 3.4 shows that the 15% threshold is exceeded by 9.7% of the observations in the error group and by 3.5% of the observations in the control group. However, the data suffer from two limitations. First, the provision of the EU-regulation applies when the 15% threshold is exceeded for more than three years²⁰⁷. To avoid losing too many observations in the sample, the analysis is limited to the cases where an auditor receives more than 15% of his revenue from a client in one year. Second, the data differ in few aspects from the requirements of the EU-regulation²⁰⁸. Due to these differences it is possible that the already few observations exceeding the 15% threshold are overestimated²⁰⁹.

²⁰³ Previous studies for Germany using the ratio of total fees to auditor's revenue include e.g.

Lopatta/Kaspereit/Canitz/Maas (2015), Quick/Sattler (2011a), Sattler (2011) and Krauß/Zülch (2013).

²⁰⁴ For a more detailed description on collecting revenue data as well as on the assumptions made to identify the relevant transparency reports, see Fiallo/Hecker (2019a), pp. 228-229 and Fiallo/Hecker (2019b), pp. 294-295.

²⁰⁵ For further details as well as references to the regulatory framework, see Fiallo/Hecker (2019b), pp. 294-295.

²⁰⁶ No observation among those excluded has a client importance higher than 15%.

²⁰⁷ For a discussion of this limitation, see Fiallo/Hecker (2019b), pp. 295-296.

²⁰⁸ The differences are discussed in Fiallo/Hecker (2019b), p. 294.

²⁰⁹ Fiallo/Hecker (2019b), p. 296.

The regression model employed for testing H3 is the following:

$$Error_{i,t} = \beta_1 * ClientImp_{i,t} + \beta_2 * RoA_{i,t} + \beta_3 * Loss_{i,t} + \beta_4 * Lev_{i,t} + \beta_5 * CFO_{i,t} + \beta_6 * Growth_{i,t} + \beta_7 * Big4_{i,t} + \beta_8 * Change_{i,t} + \beta_9 * IFRSfirst_{i,t} + \beta_{10} * Size_{i,t} + \varepsilon_{i,t}$$

3.4.4 Control variables

In the models, I control for several factors affecting the probability of error such as the company's economic situation, firm complexity, and several auditor-related characteristics²¹⁰. To account for firm performance, the variables *RoA* and *Loss* are employed. *RoA* is the return on assets and is measured as net income on total assets, while *Loss* is a dummy variable, assuming the value of 1, if the company reports a loss in year *t*, and zero otherwise. Financially distressed companies are identified by including the variables *Lev*, which is the ratio of total liabilities to total assets and *CFO*, which is operating cash-flow scaled by total assets. I expect poor performance and financial distress to increase the probability of having enforcement findings, as managers are more likely to resort to earnings management and have fewer resources to invest in high accounting quality²¹¹.

Complexity is measured by the variables *Growth* and *IFRSfirst*. Complexity increases for fast growing companies, as the accounting department and the external auditor, might struggle to keep pace. Therefore, I include the variable *Growth*, which is the change in revenues from the previous year, expecting a positive association with the existence of enforcement findings²¹². Over the years, another source of complexity has been the first-time application of IFRS. Therefore, I create the dichotomous variable *IFRSfirst*, which is coded 1 in the first year of IFRS application, and 0 in the following years. As the FREP also identifies the lack of experience with IFRS as a driver of accounting errors²¹³, I expect a positive association of *IFRSfirst* with the existence of enforcement findings.

²¹⁰ By employing variables measured on the base of erroneous financial statements, it is possible to incur in measurement errors, due to the underlying data being incorrect. However, this approach seems to be common. E.g. Gros (2016) and Markelevich/Rosner (2013) also employ accounting variables from erroneous financial statements. To avoid this issue, studies might use non-financial metrics, see e.g. the considerations made by Markelevich/Rosner (2013), p. 1608, on using as a proxy for financial distress the difference between the percentage change in the number of employees and the percentage change in total assets. Another approach is to employ variables measured in the year prior to misstatements. However, prior year financial statements might also be erroneous.

²¹¹ E.g. Marten/Quick/Ruhnke (2015), pp. 293-294; Ettredge/Scholz/Smith/Sun (2010), p. 338.

²¹² E.g. Krauß/Zülch (2013), p. 316; Gros (2016), p. 238; Sattler (2011), p. 395.

²¹³ E.g. FREP (2018a), p. 6.

The models control also for auditor-related characteristics, namely auditor size and auditor change. The variable *Big4*, which is a dichotomous variable, assumes the value of 1, if the statutory auditor is one of the Big 4, and 0 otherwise. Big audit companies tend to be associated with higher audit quality, because they might have more resources, more detailed know-how and might count on the international connections within their network²¹⁴. Additionally, big audit firms might have more to lose in terms of reputation and quasi-rents, which is why they might behave more independently²¹⁵. Therefore, I expect auditor's size to be negatively associated with the existence of enforcement findings. An auditor's change is usually advocated as a means of increasing auditor independence, but evidence shows that audit quality might decrease in the first year after an auditor's change²¹⁶. Whenever an auditor does not have previous knowledge or experience with the client, errors are more likely to occur²¹⁷. The variable *Change* is coded 1 if the auditor changed compared to the previous year, and 0 otherwise.

Finally, I employ the variable *Size*, measured as the natural log of total assets. When the matching is not "perfect", as in this case where the error company and its match are of a similar but not the same size, the difference in the matching variable can bias the results²¹⁸. Therefore, it is important to include *Size* in the analysis to account for the residual pair-wise difference²¹⁹.

²¹⁴ E.g. Krauß/Zülch (2013), p. 316.

²¹⁵ DeAngelo (1981a), p. 117; DeAngelo (1981b), p.184.

²¹⁶ E.g. Quick/Wiemann (2011), p. 934; Krauß/Zülch (2013), p. 316.

²¹⁷ E.g. Marten/Quick/Ruhnke (2015), p. 299; a similar assumption is reported by Fiallo/Hecker (2019c), p. 344, although their results do not provide evidence that audit quality is lower in the first three engagement years (pp. 345-346).

²¹⁸ Cram/Karan/Stuart (2009), pp. 479-481; Stuart/Shin/Cram/Karan (2013), p. 90. This issue has been discussed also in Fiallo/Hecker (2019a), p. 233.

²¹⁹ To account for the pair-wise difference in size, it is indicated to include the variable used in the matching process, in this case total assets. However, I prefer using the natural log of total assets since I expect size to behave non-linearly in the model. On the trade-off between fully accounting for the imperfect matching and modelling the variable's contribution, see Cram/Karan/Stuart (2009), pp. 489, 503.

Table 3.5: Variables

Variables	Description	Effect
<i>Error_{i,t}</i>	Dichotomous variable coded 1 if financial statements for the year t of company i are subject to an enforcement announcement, and zero otherwise.	-
<i>NAScat_b_{i,t}</i>	Natural log of fees disclosed by company i in year t for other assurance services.	Positive/ Negative
<i>NAScat_c_{i,t}</i>	Natural log of fees disclosed by company i in year t for tax services.	Positive/ Negative
<i>NAScat_d_{i,t}</i>	Natural log of fees disclosed by company i in year t for other non-audit services in year t.	Positive/ Negative
<i>NASratio_b_{i,t}</i>	Ratio of fees disclosed by company i in year t for other assurance services to total fees.	Positive/ Negative
<i>NASratio_c_{i,t}</i>	Ratio of fees disclosed by company i in year t for tax services to total fees.	Positive/ Negative
<i>NASratio_d_{i,t}</i>	Ratio of fees disclosed by company i in year t for other non-audit services to total fees.	Positive/ Negative
<i>AuditFees_{i,t}</i>	Natural log of fees disclosed by company i in year t for audit services.	Positive/ Negative
<i>Bilmog_{i,t}</i>	Dichotomous variable coded 1, if financial statements of company i in year t begin after 31 December 2008, and 0 otherwise.	Positive/ Negative
<i>Cap_{i,t}</i>	Dichotomous variable coded 1 if company i during t, t-1, t-2 and t-3 acquired NAS in each year, appointed the same group auditor, was a PIE, and had in year t a ratio of NAS fees to the average audit fees charged in t-1, t-2 and t-3 greater than 70%, and 0 otherwise.	Positive/ Negative
<i>ClientImp_{i,t}</i>	Dichotomous variable coded 1, if the ratio of total fees disclosed by company i in year t to the auditor's revenue disclosed in t+1 exceeds 15%, and zero otherwise.	Positive
<i>RoA_{i,t}</i>	Ratio of net income to total assets for company i in year t.	Negative
<i>Loss_{i,t}</i>	Dichotomous variable coded 1, if company i reports a loss in year t, and zero otherwise.	Positive
<i>Lev_{i,t}</i>	Ratio of total liabilities to total assets for company i in year t.	Positive
<i>CFO_{i,t}</i>	Ratio of operating cash-flow to total assets for company i in year t.	Negative
<i>Growth_{i,t}</i>	Percentage change in revenues for company i in year t compared to t-1.	Positive
<i>IFRSfirst_{i,t}</i>	Dichotomous variable coded 1, if company i in year t applied IFRS for the first time, and zero otherwise.	Positive
<i>Big4_{i,t}</i>	Dichotomous variable coded 1, if the group auditor of company i in year t is one of the Big 4, and zero otherwise.	Negative
<i>Change_{i,t}</i>	Dichotomous variable coded 1, if the group auditor of company i in year t is appointed for the first time, and zero otherwise.	Positive
<i>Size_{i,t}</i>	Natural log of total assets for company i in year t.	Positive/ Negative

Notes: The expected effect is estimated referring to the probability of having enforcement findings. Data underlying the following variables have been manually collected from consolidated financial statements and auditor's transparency reports: *NAScat_b_{i,t}*, *NAScat_c_{i,t}*, *NAScat_d_{i,t}*, *NASratio_b_{i,t}*, *NASratio_c_{i,t}*, *NASratio_d_{i,t}*, *AuditFees_{i,t}*, *Bilmog_{i,t}*, *Cap_{i,t}*, *ClientImp_{i,t}*, *IFRSfirst_{i,t}*, *Big4_{i,t}*, *Change_{i,t}*. All other underlying data have been retrieved from Thomson Reuters Eikon, where they are available under the following codes: TR.Revenue, TR.TotalLiabilities, TR.TotalAssets, TR.NetIncomeBeforeExtraItems, TR.CashFromOperatingAct.

3.5 Results

3.5.1 Descriptive statistics

When performing univariate tests, it is important to account for paired observations²²⁰. Due to the “imperfect” matching, popular tests like the one-sample t-test for matched data may not be appropriate, as the residual size difference, rather than the group membership, could drive the results²²¹. Therefore, I run separated regressions for each variable including dummies for pairs and using the existence of enforcement findings and size as explanatory variables²²². I control for extreme values, by winsorizing all continuous variables at the 1% and 99% level.

Table 3.6 shows that companies with enforcement findings buy on average fewer tax services than companies in the control group. The coefficients of *NAScat_c* and *NASratio_c* are both significant at the 1% level. The average of *NAScat_d* is also lower for the error group (p-value < 0.01), which might suggest that error companies pay lower fees for other non-audit services. However, there is no significant association between *Error* and *NASratio_d*. Regarding other assurance services, there is only weak evidence that error companies buy more of this type of services compared to the control group (*NASratio_b*, p-value < 0.1). These results are comparable to the descriptive results obtained, following different approaches, by *Fiallo/Hecker* (2019a)²²³. Their analysis of frequency shows that there are more companies in the control group buying tax and other non-audit services than in the error group, while there are slightly more companies buying other assurance services in the error group (44.8%) than in the control group (44.1%). Also, the univariate conditional logistic regression shows that companies buying tax services and other non-audit services are less likely to have enforcement findings, while regarding other assurance services, there is no significant difference among the two groups.

²²⁰ A similar consideration is made in *Fiallo/Hecker* (2019a), p. 233 footnote no. 76.

²²¹ *Cram/Karan/Stuart* (2009), pp. 494-495.

²²² As suggested by *Cram/Karan/Stuart* (2009), pp. 494-495, I run the regression including an overall intercept and omitting one dummy pair. As reported by *Cram/Karan/Stuart* (2009), p. 495 the results of an OLS on choice-based data cannot be generalised as “there exists no larger population to which the sample is similar”. The results should be corrected by reweighting the OLS by the inverse of the sampling rate in each stratum. The strata in the sample are, among others, specific to industry, year and accounting standard, which would make the collection of sampling information costly to implement. By not reweighting the OLS, I caution that the results are valid only for the analysed sample. Conversely, when running a conditional logistic regression, this problem does not occur. See *Cram/Karan/Stuart* (2009), pp. 492-495.

²²³ *Fiallo/Hecker* (2019a), pp. 232-234.

Table 3.6: Descriptive statistics

Variables	Error group						Control group						OLS/Clogit regression	
	Mean	Median	St. dev.	Min.	Max.	Obs. (m)	Mean	Median	St. dev.	Min.	Max.	Obs. (m)	coeff.	p-value
<i>Error</i>	1.000	1.000	0.000	1.000	1.000	141	0.000	0.000	0.000	0.000	0.000	141		
<i>NAScat_b</i>	5.031	0.000	5.820	0.000	15.464	141	4.800	0.000	5.551	0.000	15.464	141	0.197	(0.751)
<i>NAScat_c</i>	4.311	0.000	5.424	0.000	14.346	141	6.609	9.210	5.293	0.000	14.346	141	-2.303***	(0.000)
<i>NAScat_d</i>	5.932	8.556	5.651	0.000	15.342	141	7.491	9.473	4.923	0.000	15.202	141	-1.591***	(0.005)
<i>NASratio_b</i>	0.094	0.000	0.151	0.000	0.638	141	0.067	0.000	0.114	0.000	0.638	141	0.027*	(0.090)
<i>NASratio_c</i>	0.062	0.000	0.112	0.000	0.519	141	0.108	0.063	0.127	0.000	0.519	141	-0.044***	(0.003)
<i>NASratio_d</i>	0.089	0.019	0.126	0.000	0.571	141	0.100	0.057	0.124	0.000	0.571	141	-0.010	(0.471)
<i>AuditFees</i>	12.314	12.196	1.297	10.127	16.232	141	12.205	11.891	1.194	10.127	16.232	141	0.092	(0.129)
<i>Cap</i>	0.111	0.000	0.316	0.000	1.000	72	0.194	0.000	0.399	0.000	1.000	72	-0.884	(0.137)
<i>ClientImp</i>	0.097	0.000	0.298	0.000	1.000	113	0.035	0.000	0.186	0.000	1.000	113	1.298*	(0.063)
<i>RoA</i>	-0.023	0.007	0.141	-0.481	0.280	141	0.020	0.026	0.111	-0.481	0.280	141	-0.043***	(0.005)
<i>Loss</i>	0.404	0.000	0.492	0.000	1.000	141	0.241	0.000	0.429	0.000	1.000	141	0.797***	(0.004)
<i>Lev</i>	0.672	0.670	0.284	0.097	1.832	141	0.572	0.592	0.286	0.081	1.832	141	0.106***	(0.001)
<i>CFO</i>	-0.002	0.025	0.142	-0.512	0.296	141	0.053	0.061	0.113	-0.427	0.296	141	-0.058***	(0.000)
<i>Growth</i>	0.349	0.066	1.408	-0.783	9.173	141	0.155	0.042	0.856	-0.783	9.173	141	0.194	(0.165)
<i>Big4</i>	0.482	0.000	0.501	0.000	1.000	141	0.652	1.000	0.478	0.000	1.000	141	-0.826***	(0.003)
<i>Change</i>	0.128	0.000	0.335	0.000	1.000	141	0.078	0.000	0.269	0.000	1.000	141	0.616	(0.138)
<i>IFRSfirst</i>	0.092	0.000	0.290	0.000	1.000	141	0.078	0.000	0.269	0.000	1.000	141	0.167	(0.737)
<i>Size</i>	19.334	18.780	2.459	14.945	26.198	141	19.294	18.902	2.304	14.945	26.198	141	0.040	(0.437)

Notes: The model for the OLS is the following: $Var_{i,t} = \beta_0 + \beta_1 Error_{i,t} + \beta_2 Size_{i,t} + \sum_{y=1}^{m-1} \beta_y Pair_y$. For the dichotomous variables *Cap*, *ClientImp*, *Loss*, *Big4*, *Change* and *IFRSfirst* I run a conditional logistic regression with pairs identified as strata. The reported coefficients are for the variable *Error*. The reported p-values are two-sided. ***, **, * means significance at the 1%, 5% and 10% level.

The coefficient of *ClientImp* shows weak evidence that auditors of companies in the control group are less likely to obtain more than 15% of their revenues from the audited entity compared to auditors of error companies. However, the coefficient is only significant at the 10% level. Similarly, the analysis in *Fiallo/Hecker (2019b)* provides only some indications that companies with a fee-to-revenue ratio higher than 15% are more likely to have enforcement findings²²⁴. The coefficient of the variable *Cap* is not significant, not delivering any indication that exceeding the cap is associated with the existence of enforcement findings. Similarly, the analysis of *Fiallo/Hecker (2019b)* does not provide any significant result²²⁵. Error and control companies differ in the sample in terms of economic performance. Error companies are less profitable and are more likely to report a loss in year t (*RoA* and *Loss*, p-values < 0.01). Additionally, they are more leveraged and have

²²⁴ Fiallo/Hecker (2019b), pp. 295-296. The coefficient of the variable of interest is significant at the 10% level.

²²⁵ Fiallo/Hecker (2019b), p. 293.

less cash-flow available, which might indicate financial distress (*Lev* and *CFO*, p-values < 0.01). Finally, they are less likely to be audited by one of the Big 4 (*Big4*, p-value < 0.01).

Table 3.7: Pearson's and Spearman's correlation coefficients (part 1)

	<i>Error</i>	<i>NAScat_b</i>	<i>NAScat_c</i>	<i>NAScat_d</i>	<i>NASratio_b</i>	<i>NASratio_c</i>	<i>NASratio_d</i>	<i>AuditFees</i>	<i>Cap</i>	<i>ClientImp</i>
<i>Error</i>		0.0603	-0.1279	-0.1745*	0.0884	-0.1713*	-0.2034*	0.0531	-0.1108	0.1113
<i>NAScat_b</i>	0.0204		0.1409	0.3857*	0.9702*	-0.0130	0.1615	0.5571*	0.2057*	0.2268*
<i>NAScat_c</i>	-0.2104*	0.1390*		0.2958*	0.0498	0.9299*	0.1983*	0.2998*	0.3544*	0.1004
<i>NAScat_d</i>	-0.1460*	0.2640*	0.2421*		0.2995*	0.1481	0.9016*	0.6003*	0.3842*	0.0754
<i>NASratio_b</i>	0.1032	0.7489*	-0.0212	0.0958		-0.0795	0.1169	0.4393*	0.1664*	0.2208*
<i>NASratio_c</i>	-0.1873*	0.0081	0.7418*	0.0833	-0.1216*		0.1413	0.0665	0.3465*	0.0668
<i>NASratio_d</i>	-0.0416	0.0001	0.0503	0.6856*	-0.0665	-0.0122		0.3077*	0.3834*	0.0240
<i>AuditFees</i>	0.0438	0.4760*	0.2313*	0.4585*	0.2083*	-0.0406	0.1460*		0.1655*	0.1418
<i>Cap</i>	-0.1158	0.1853*	0.3079*	0.3312*	0.1645*	0.4287*	0.4473*	0.1626		0.0214
<i>ClientImp</i>	0.1244	0.0051	-0.1716*	-0.1659*	0.1581*	-0.1713*	-0.1546*	-0.1255	-0.0991	
<i>RoA</i>	-0.1653*	0.1109	0.1290*	0.0929	0.0404	0.1125	0.0153	0.1364*	0.0507	-0.0892
<i>Loss</i>	0.1745*	-0.0549	-0.1197*	-0.0707	0.0068	-0.1285*	-0.0740	-0.1100	-0.0498	0.1141
<i>Lev</i>	0.1746*	0.0812	-0.0189	0.1747*	0.0687	-0.0773	0.0755	0.2634*	0.0487	-0.0596
<i>CFO</i>	-0.2094*	0.1044	0.0504	0.0193	-0.0078	-0.0056	-0.0740	0.2043*	0.1273	-0.1817*
<i>Growth</i>	0.0834	-0.0184	0.0073	-0.0027	0.0928	-0.0441	-0.0067	-0.1075	-0.0610	0.1846*
<i>Big4</i>	-0.1718*	0.2189*	0.0957	0.2590*	0.1060	-0.0078	0.1279*	0.4395*	0.1012	-0.3434*
<i>Change</i>	0.0817	-0.1529*	-0.0995	-0.1234*	-0.1162	-0.0754	-0.1059	-0.0756	-0.1448	0.0760
<i>IFRSfirst</i>	0.0254	-0.0722	-0.0628	-0.0693	-0.0697	-0.0375	-0.0693	-0.0798	0.0457	0.0366
<i>Size</i>	0.0085	0.4571*	0.1758*	0.4713*	0.2011*	-0.0690	0.2250*	0.8968*	0.1710*	-0.1164

Notes: The table reports the coefficients of the Pearson product-moment correlation (under the diagonal) and the Spearman rank-order correlation (above the diagonal). * means significance at 5% level or lower.

Tables 3.7 and 3.8 show the Pearson's and Spearman's correlation coefficients²²⁶. In line with the findings obtained so far, the variable *Error* is negatively correlated with the fees paid for tax services (*NAScat_c*, *NASratio_c*) and for other non-audit services (*NAScat_d*). Also, *Error* is negatively correlated with performance (*RoA*) and liquidity (*CFO*), while it is positively correlated with leverage (*Lev*) and *Loss*²²⁷. Also, Tables 3.7 and 3.8 show that *Loss* and *RoA* are strongly negatively correlated (-0.67) and that *ClientImp* is strongly negatively associated with *Big4* (-0.68). This is not surprising, as due to their broad portfolio of clients and to their size, it is unlikely for Big 4 auditors to receive more than 15% of their revenues from a client. As expected, strong

²²⁶ The Pearson's correlation coefficients can be interpreted even when one of the two variables is a category coded 0/1. This special case is known as point-biserial correlation. See e.g. Sheskin (2000), pp. 799-804, and Corder/Foreman (2009), pp. 134-142. In general, Pearson's correlation requires a linear relationship between the variables and for them to be numerical. Therefore, in Tables 3.7 and 3.8 are also reported the Spearman's rank correlation coefficients, which can be calculated for ordinal variables. See e.g. Cleff (2019), pp. 94-95.

²²⁷ *Loss* is significant at the 5% level according to the Spearman's rank.

correlations (> 0.5) exist between *NAScat_b* and *NASratio_b* (0.7489), between *NAScat_c* and *NASratio_c* (0.7418), between *NAScat_d* and *NASratio_d* (0.6856). The two sets of fee variables, namely the logarithmic transformations of fees and the fee ratios, are not used together in any model. Also, the correlation coefficients show that bigger companies pay higher fees for audit services, as indicated by the significant positive correlation between *AuditFees* and *Size* (0.8968). The variables *AuditFees* and *Size* are the logarithmic transformation of fees for audit services and of total assets, which are also strongly correlated²²⁸. Due to the existence of strong correlations between several variables, I further calculate the Variance Inflation Factor (VIF) to exclude the existence of multicollinearity²²⁹. The results (untabulated) show that all VIFs are below the critical value of 10, indicating that the model should not suffer from multicollinearity issues²³⁰.

Table 3.8: Pearson's and Spearman's correlation coefficients (part 2)

	<i>RoA</i>	<i>Loss</i>	<i>Lev</i>	<i>CFO</i>	<i>Growth</i>	<i>Big4</i>	<i>Change</i>	<i>IFRSfirst</i>	<i>Size</i>
<i>Error</i>	-0.1368	0.1710*	0.1881*	-0.1311	0.0866	-0.0981	0.0047	0.0024	0.0148
<i>NAScat_b</i>	-0.0543	0.0549	0.3272*	0.0390	-0.0165	0.2148*	-0.0894	-0.0189	0.5238*
<i>NAScat_c</i>	0.1382	-0.1160	0.0339	0.0886	-0.0426	0.1332	-0.0844	-0.0991	0.2671*
<i>NAScat_d</i>	0.0709	-0.1171	0.2820*	0.0297	-0.0674	0.3397*	-0.0984	-0.0945	0.6398*
<i>NASratio_b</i>	-0.0666	0.0776	0.2714*	0.0276	0.0195	0.1427	-0.0973	-0.0023	0.4075*
<i>NASratio_c</i>	0.1972*	-0.1434	-0.0675	0.0369	-0.0552	0.0097	-0.0717	-0.0807	0.0468
<i>NASratio_d</i>	0.1317	-0.1566	0.1319	0.0239	-0.0270	0.1969*	-0.0670	-0.0710	0.3963*
<i>AuditFees</i>	0.0033	-0.0622	0.4231*	0.0378	-0.1469	0.4429*	0.0009	-0.1739*	0.9037*
<i>Cap</i>	0.1130	-0.0487	0.0537	0.1614	-0.0358	0.0949	-0.1416	0.0447	0.1956*
<i>ClientImp</i>	-0.0378	0.0313	0.1635	-0.1474	-0.0369	-0.6795*	0.0484	-0.0696	0.1310
<i>RoA</i>		-0.8378*	-0.3163*	0.4395*	0.2627*	-0.0782	-0.0905	-0.0405	0.0766
<i>Loss</i>	-0.6739*		0.2330*	-0.2970*	-0.2071*	0.0346	0.0378	0.0446	-0.1572
<i>Lev</i>	-0.3832*	0.2294*		-0.2988*	-0.0592	0.1297	-0.0807	-0.0291	0.3395*
<i>CFO</i>	0.4482*	-0.3687*	-0.2901*		0.0373	0.0802	-0.0392	-0.0644	0.0358
<i>Growth</i>	0.1037	-0.0921	-0.0431	-0.0592		-0.0612	-0.1153	0.1173	-0.0709
<i>Big4</i>	0.0193	-0.0097	0.0204	0.1546*	-0.1126		-0.0816	-0.0420	0.4332*
<i>Change</i>	-0.0712	0.0909	-0.0656	-0.0532	-0.0661	-0.0578		-0.0563	0.0594
<i>IFRSfirst</i>	-0.0057	-0.0474	0.1030	-0.0752	0.1667*	-0.1184*	-0.0614		-0.1776*
<i>Size</i>	0.2010*	-0.1816*	0.1855*	0.1914*	-0.0859	0.4142*	-0.0374	-0.0584	

Notes: The table reports the coefficients of the Pearson product-moment correlation (under the diagonal) and the Spearman rank-order correlation (above the diagonal). * means significance at 5% level or lower.

²²⁸ The Pearson's correlation coefficient of audit fees and total assets is 0.882.

²²⁹ Wooldridge (2016), p. 86.

²³⁰ Other studies, which have used 10 as a cut-off value, are e.g. Lopatta/Kaspereit/Canitz/Maas (2015), p. 580; Gros (2016), p. 235.

3.5.2 Multivariate results

3.5.2.1 H1: NAS categories

To test H1, I run a conditional logistic regression on the models M1-M4 using alternatively the natural log of NAS fees from category b), c) and d) (*NAScat_b*, *NAScat_c* and *NAScat_d*) and the ratios of the fees disclosed in each category to total fees (*NASratio_b*, *NASratio_c* and *NASratio_d*). The results presented in Tables 3.9 and 3.10 show that according to the X^2 test, the models are all statistically significant (p-values < 0.01) and they have an explanatory power between 21% and 36%, which is in line with prior studies²³¹.

The coefficients of *NAScat_b* and *NASratio_b* are both not significant in M1 and M2, indicating the lack of association between fees for other assurance services and enforcement findings. In this case, there is no evidence that the provision of other assurance services is associated with audit quality. The coefficients of *NAScat_c* and *NASratio_c* are both significant and negatively associated with the dependent variable *Error* in models M1 and M3 (p-values < 0.05). These results suggest that the provision of tax services might generate knowledge spillovers, which have a positive impact on audit quality. It is possible that certain tax services, such as the annual tax declaration, are particularly suitable for knowledge spillovers due to their closeness to the related tax accounts in the financial statements²³². Regarding other non-audit services, the results are mixed: while the coefficient of *NAScat_d* is negatively associated with enforcement findings in both models M1 and M4 (p-value < 0.01), the coefficient of *NASratio_d* is not significant. The results deliver only partial evidence that other non-audit services are positively associated with audit quality.

In models M1 through M4, *AuditFees* is significant and positively associated with the existence of enforcement findings²³³. I do not predict the sign of the coefficient of *AuditFees*, as its impact on audit quality is unclear. On the one hand, the level of audit fees should reflect the auditor's effort to reduce audit risk to an acceptable level²³⁴, on the other hand, high audit fees might increase the

²³¹ E.g. Gros (2016), p. 237.

²³² Francis (2004), p. 363.

²³³ In M2 the coefficient is only significant at the 10% level. In M4 using the fee ratios, the coefficient of *AuditFees* is also significant at the 10% level.

²³⁴ Simunic (1980), p. 166.

quasi-rents and impair auditor independence²³⁵. The results show that high audit fees are associated with lower audit quality, which might indicate the existence of an economic bond²³⁶.

Regarding the measures of performance and financial distress, the dependent variable *Error* is positively correlated with *Loss* in M2 and M3, even though the coefficients are weakly significant (p-values < 0.10) and is negatively associated with *CFO* in each model (p-value < 0.01). The results provide only partial evidence that loss-making companies are more likely to have enforcement findings but suggest that companies with a positive cash balance from operating activities are less likely to be subject to an enforcement announcement. Conversely, the coefficients of the other measures of performance and financial distress (*RoA* and *Lev*) are not significant throughout the models. Thus, it cannot be concluded whether poor performance and financial distress are associated with low audit quality.

The coefficient of the variable *Big4* is negative and significant in each model (p-values < 0.01). This indicates an inverse relationship between auditor size and the presence of enforcement findings, which suggests that Big 4 auditors might deliver higher quality. The quasi-rent theory of *DeAngelo* (1981b) affirms that auditor independence depends on the cost-benefit consideration of keeping a client against the probability of being discovered “cheating” and losing other clients²³⁷. It is possible that as usually big audit firms have a broader portfolio of clients, they might behave more independently²³⁸.

A possible confounding factor, which has not been considered until now, is that the disclosure requirements have changed with the introduction of BilMoG²³⁹. With valuation services being reclassified from category b) to d), the variables for the two categories might not measure the same thing over time²⁴⁰. The variable *Bilmog* is added to the models M2 and M4 to test whether there is a statistical interaction between *Bilmog*, and the fees disclosed under category b) and category d)

²³⁵ DeAngelo (1981a), p. 113.

²³⁶ Fiallo/Hecker (2019a), pp. 233-234, find that companies buying only audit services are more likely to have enforcement findings.

²³⁷ DeAngelo (1981b), p. 197. See also the discussion in section 3.3.1.2.

²³⁸ DeAngelo (1981a), p. 117; DeAngelo (1981b), p.184.

²³⁹ This issue is discussed in section 3.4.2. For a description of the changes in the disclosure of auditor’s fees introduced by BilMoG, see Fiallo/Hecker (2019a), pp. 229-230.

²⁴⁰ This is not the only change brought by BilMoG, see here e.g. Fiallo/Hecker (2019a), pp. 229-230. However, in this study I focus on this issue as I deem it to be particularly relevant when trying to answer the question of whether other assurance services and other non-audit services are associated with the existence of enforcement announcements.

(*NAScat_b*, *NASratio_b*, *NAScat_d*, *NASratio_d*)²⁴¹. In other words, the analysis aims at assessing whether the odds of having enforcement findings are more sensitive to services from the categories b) and d) before or after BilMoG²⁴².

As shown in Table 3.11, the coefficient of *NAScat_d* is negative and significant (p-value < 0.5). This means that there is a negative association between *NAScat_d* and *Error* when *Bilmog* is set to zero, indicating that in the period before BilMoG the odds of a company being subject to enforcement findings, when purchasing high levels of other non-audit services, are smaller (0.9) compared to the period after BilMoG. Similarly, the sign and magnitude of the coefficient of the interaction term might indicate that the effect of *NAScat_d* on the dependent variable is stronger (more negative) after BilMoG, although, it is not significant. No evidence is provided by the other models, where *NAScat_b*, *NASratio_b*, *NASratio_d* and *Bilmog* are all not significant. It follows that there is not enough evidence to answer the question whether the odds of having enforcement findings are more sensitive to services from the categories b) and d) before or after BilMoG. Also, as the interaction term is not significant in any model, this might indicate that the main models as reported in Tables 3.9 and 3.10 do not omit a statistical interaction.

This analysis suffers from at least two issues²⁴³. First, the classification of the provided services in categories, as reported in financial statements, might differ in some respects from the definition of audit services, other assurance services, tax services and other non-audit services as intended by art. 5 EU-reg. Second, by design, companies can be included multiple times in the sample. It follows that the underlying variables may not be independent.

²⁴¹ For an analysis of statistical interaction in nonlinear models, see Hosmer/Lemeshow/Sturdivant (2013), pp. 64-77. On the importance of adding all constitutive terms of the interaction in the model, see Brambor/Clark/Golder (2006), pp. 66-70; 77; Hosmer/Lemeshow/Sturdivant (2013), pp. 73. 77.

²⁴² As to the why I talk about odds and not likelihood, see Powers (2005), p. 505; Buis (2010), pp. 305-306.

²⁴³ For a discussion of the limitations see Fiallo/Hecker (2019a), p. 231.

Table 3.9: Multivariate results for H1- part 1

	expected sign	(M1)	(M1)	(M2)	(M2)	(M3)	(M3)	(M4)	(M4)
		coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio
<i>NAScat_b</i>	-/+	-0.008 (0.811)	0.992	0.004 (0.891)	1.004				
<i>NAScat_c</i>	-/+	-0.120*** (0.000)	0.887***			-0.120*** (0.000)	0.887***		
<i>NAScat_d</i>	-/+	-0.123*** (0.001)	0.884***					-0.128*** (0.001)	0.880***
<i>AuditFees</i>	-/+	1.223*** (0.002)	3.398***	0.576* (0.063)	1.780*	1.018*** (0.005)	2.767***	0.752** (0.024)	2.122**
<i>RoA</i>	-	0.335 (0.566)	1.398	0.137 (0.529)	1.147	0.232 (0.548)	1.262	0.274 (0.557)	1.316
<i>Loss</i>	+	0.556 (0.131)	1.744	0.672* (0.061)	1.958*	0.666* (0.075)	1.947*	0.563 (0.110)	1.756
<i>Lev</i>	+	0.465 (0.276)	1.593	0.734 (0.143)	2.083	0.159 (0.414)	1.172	0.918 (0.102)	2.504
<i>CFO</i>	-	-5.976*** (0.000)	0.003***	-3.472*** (0.009)	0.031***	-4.468*** (0.002)	0.011***	-5.088*** (0.001)	0.006***
<i>Growth</i>	+	0.190 (0.113)	1.210	0.149 (0.145)	1.161	0.133 (0.186)	1.142	0.185 (0.105)	1.203
<i>Big4</i>	-	-1.367*** (0.001)	0.255***	-1.170*** (0.000)	0.310***	-1.323*** (0.000)	0.266***	-1.250*** (0.000)	0.287***
<i>Change</i>	+	0.193 (0.374)	1.212	0.381 (0.216)	1.463	0.395 (0.231)	1.484	0.201 (0.355)	1.223
<i>IFRSfirst</i>	+	-0.437 (0.766)	0.646	-0.161 (0.621)	0.851	-0.177 (0.623)	0.838	-0.493 (0.811)	0.611
<i>Size</i>	-/+	0.233 (0.582)	1.263	0.257 (0.476)	1.293	0.060 (0.876)	1.062	0.394 (0.292)	1.483
Obs.		282		282		282		282	
Prob > X ²		0.000		0.000		0.000		0.000	
Pseudo R ²		0.356		0.210		0.295		0.281	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff.}$ and indicate the percentage change in the probability of the outcome occurring (the company having an enforcement announcement) versus the outcome not occurring, by a 1-unit increase in the independent variable. In the case of dichotomous predictors, they indicate the increase or decrease in the odds by the occurrence of the effect coded in the categorical variable. For more considerations on odds ratios, see Hosmer/Lemeshow/Sturdivant (2013), pp. 50-52, 63. Prob > X² indicates at which level it is possible to reject the null hypothesis that the predictors have no effect on the dependent variable. The goodness of fit is expressed by the McFadden's pseudo R².

Table 3.10: Multivariate results for H1 - part 2

	expected sign	(M1)	(M1)	(M2)	(M2)	(M3)	(M3)	(M4)	(M4)
		coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio
<i>NASratio_b</i>	-/+	0.857 (0.477)	2.355	1.421 (0.221)	4.141				
<i>NASratio_c</i>	-/+	-3.466** (0.011)	0.031**			-3.500*** (0.009)	0.030***		
<i>NASratio_d</i>	-/+	-1.241 (0.347)	0.289					-0.805 (0.522)	0.447
<i>AuditFees</i>	-/+	0.693** (0.038)	2.000**	0.564* (0.068)	1.757*	0.736** (0.025)	2.088**	0.554* (0.075)	1.741*
<i>RoA</i>	-	-0.157 (0.467)	0.854	-0.123 (0.474)	0.884	0.209 (0.544)	1.232	0.014 (0.503)	1.014
<i>Loss</i>	+	0.457 (0.166)	1.579	0.615* (0.082)	1.849*	0.608* (0.087)	1.837*	0.608* (0.085)	1.837*
<i>Lev</i>	+	0.462 (0.261)	1.587	0.756 (0.137)	2.130	0.435 (0.270)	1.545	0.725 (0.145)	2.065
<i>CFO</i>	-	-4.197*** (0.005)	0.015***	-3.224** (0.014)	0.040**	-4.016*** (0.004)	0.018***	-3.662*** (0.007)	0.026***
<i>Growth</i>	+	0.095 (0.261)	1.099	0.122 (0.196)	1.130	0.108 (0.227)	1.114	0.149 (0.148)	1.161
<i>Big4</i>	-	-1.252*** (0.000)	0.286***	-1.225*** (0.000)	0.294***	-1.216*** (0.000)	0.297***	-1.168*** (0.000)	0.311***
<i>Change</i>	+	0.420 (0.205)	1.521	0.433 (0.190)	1.542	0.421 (0.199)	1.523	0.347 (0.237)	1.415
<i>IFRSfirst</i>	+	-0.162 (0.618)	0.851	-0.085 (0.564)	0.919	-0.133 (0.598)	0.876	-0.206 (0.651)	0.814
<i>Size</i>	-/+	0.032 (0.931)	1.033	0.220 (0.543)	1.246	0.011 (0.976)	1.011	0.282 (0.434)	1.326
Obs.		282		282		282		282	
Prob > X ²		0.000		0.000		0.000		0.000	
Pseudo R ²		0.256		0.217		0.249		0.212	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff.}$ and indicate the percentage change in the probability of the outcome occurring (the company having an enforcement announcement) versus the outcome not occurring, by a 1-unit increase in the independent variable. In the case of dichotomous predictors, they indicate the increase or decrease in the odds by the occurrence of the effect coded in the categorical variable. For more considerations on odds ratios, see Hosmer/Lemeshow/Sturdivant (2013), pp. 50-52, 63. Prob > X² indicates at which level it is possible to reject the null hypothesis that the predictors have no effect on the dependent variable. The goodness of fit is expressed by the McFadden's pseudo R².

Table 3.11: Multivariate results for H1 - with interaction term

	expected sign	(M2)	(M2)	(M4)	(M4)	(M2)	(M2)	(M4)	(M4)
		coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio
<i>NAScat_b</i>	-/+	-0.003 (0.934)	0.997						
<i>NAScat_d</i>	-/+			-0.105** (0.034)	0.900**				
<i>NASratio_b</i>	-/+					1.619 (0.294)	5.046		
<i>NASratio_d</i>	-/+							0.021 (0.989)	1.021
<i>Bilmog</i>	-/+	1.374 (0.353)	3.952	2.131 (0.258)	8.423	1.464 (0.313)	4.322	1.770 (0.237)	5.869
<i>Interaction</i>	-/+	0.019 (0.737)	1.019	-0.055 (0.469)	0.947	-0.529 (0.823)	0.589	-2.797 (0.300)	0.061
<i>AuditFees</i>	-/+	0.610* (0.052)	1.840*	0.788** (0.019)	2.198**	0.605* (0.054)	1.832*	0.580* (0.063)	1.787*
<i>RoA</i>	-	0.284 (0.560)	1.328	0.275 (0.556)	1.317	0.009 (0.502)	1.009	0.158 (0.533)	1.172
<i>Loss</i>	+	0.681* (0.060)	1.975*	0.557 (0.113)	1.746	0.643* (0.075)	1.901*	0.597* (0.092)	1.816*
<i>Lev</i>	+	0.754 (0.138)	2.126	0.998* (0.086)	2.714*	0.791 (0.128)	2.205	0.738 (0.144)	2.091
<i>CFO</i>	-	-3.524*** (0.009)	0.029***	-4.885*** (0.001)	0.008***	-3.243** (0.015)	0.039**	-3.704*** (0.007)	0.025***
<i>Growth</i>	+	0.151 (0.145)	1.163	0.179 (0.115)	1.196	0.119 (0.205)	1.126	0.158 (0.134)	1.171
<i>Big4</i>	-	-1.198*** (0.000)	0.302***	-1.296*** (0.000)	0.274***	-1.269*** (0.000)	0.281***	-1.212*** (0.000)	0.298***
<i>Change</i>	+	0.452 (0.178)	1.571	0.258 (0.320)	1.294	0.494 (0.164)	1.639	0.428 (0.194)	1.535
<i>IFRSfirst</i>	+	-0.194 (0.641)	0.824	-0.454 (0.792)	0.635	-0.057 (0.542)	0.944	-0.146 (0.606)	0.864
<i>Size</i>	-/+	0.254 (0.484)	1.289	0.326 (0.403)	1.385	0.214 (0.556)	1.239	0.227 (0.536)	1.255
Obs.		282		282		282		282	
Prob > X ²		0.000		0.000		0.000		0.000	
Pseudo R ²		0.216		0.290		0.223		0.223	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. *Interaction* is the interaction term between *Bilmog* and each of the following variables: *NAScat_b*, *NAScat_d*, *NASratio_b*, and *NASratio_d*. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff.}$.

3.5.2.2 H2: the cap

To test H2, I run a conditional logistic regression using the variable *Cap*. The results are presented in Table 3.12. According to the X^2 test, the model is weakly statistically significant (p-value < 0.1) and has low explanatory power. The results show a weak positive association between *Loss* and *Error* (p-value < 0.1). Additionally, the negative coefficient of the variable *Big4* shows an inverse relationship between auditor size and the presence of enforcement findings.

The coefficient of *Cap* is negative and not significant, providing no indications that the variable is associated with the existence of enforcement findings²⁴⁴. It follows that there is no evidence supporting the introduction of the cap as a mean to improve auditor independence and audit quality. Also, prior studies suggest a limited impact of the cap among PIEs²⁴⁵. Similarly, I observe that only 15.3 % of the companies in the sample have fees above the cap, while all other companies have fee levels below the 70% threshold or do not qualify for its application. *Fiallo/Hecker* (2019b) call attention to the fact that in the analysed period, the black list is not yet implemented, which means the potential impact of the cap might be even smaller once the provisions of art. 5 EU-reg. apply²⁴⁶. Given its potentially limited application as well as the lack of evidence supporting the choice of the 70% threshold, it remains unclear if and to what extent audit quality might benefit from the introduction of the cap.

This analysis suffers from some issues²⁴⁷. First, the available data allow only to approximate the calculation of the cap prescribed by the EU-regulation, as the fees collected from financial statements and used to calculate the variable *Cap* differ in some respects from the fees required by art. 4 para. 2 EU-reg. Second, the calculation of the variable *Cap* requires 4 years of fee data. If during this period fees are not disclosed consistently by the companies²⁴⁸, this might introduce a measurement error. Third, the sample design, which allows a company to be included multiple times, might be a source of bias²⁴⁹. Fourth, the lack of significant results might be due to the small

²⁴⁴ This finding is in line with that of *Fiallo/Hecker* (2019b), pp. 292-293.

²⁴⁵ E.g. *Ratzinger-Sakel/Schönberger* (2015), p. 62; *Fiallo/Hecker* (2019b), pp. 292-293.

²⁴⁶ *Fiallo/Hecker* (2019b), pp. 292-293.

²⁴⁷ For a discussion of the limitations see *Fiallo/Hecker* (2019b), pp. 291-293.

²⁴⁸ E.g., due to general uncertainties on the part of the companies regarding the interpretation of the disclosure requirements.

²⁴⁹ In the sample used to test H2, a company is repeated a maximum of two times. Also, as reported in *Fiallo/Hecker* (2019b), p. 293 footnote no. 105, due to the 1:1 match, some observations are dropped due to missing data for the matched company. Given the already small occurrence of cases where fees exceed the cap, it should be mentioned that 5 observations which are dropped from the error group have *Cap* = 1.

sample size, as well as the relative low occurrence of cases where *Cap* was equal to 1 (8 times in the error group and 14 in the control group)²⁵⁰.

Table 3.12: Multivariate results for H2

	expected sign	coeff. (p-value)	odds ratio
<i>Cap</i>	-/+	-0.827 (0.148)	0.437
<i>RoA</i>	-	1.806 (0.756)	6.085
<i>Loss</i>	+	0.934* (0.052)	2.545*
<i>Lev</i>	+	1.048 (0.115)	2.853
<i>CFO</i>	-	-2.020 (0.159)	0.133
<i>Growth</i>	+	-0.163 (0.715)	0.850
<i>Big4</i>	-	-0.869** (0.030)	0.419**
<i>Change</i>	+	-0.499 (0.780)	0.607
<i>IFRSfirst</i>	+	0.666 (0.309)	1.947
<i>Size</i>	-/+	0.965 (0.105)	2.624
Obs.		144	
Prob > X ²		0.096	
Pseudo R ²		0.161	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff}$. The analysis is based on a restricted sample of 144 observations.

²⁵⁰ As emphasised by Hosmer/Lemeshow/Sturdivant (2013), p. 246, the explanatory power of the conditional model comes from discordant pairs. In this analysis, there are only 18 discordant pairs.

3.5.2.3 H3: Client importance

To test H3, I employ the variable *ClientImp*. The results are presented in Table 3.13. The reported X^2 test indicates that the model is statistically significant (p-values < 0.01) but has low explanatory power (pseudo $R^2 = 0.16$). The results show a weak positive association between *Lev* and *Error* and a negative association between *CFO* and *Error*, which might indicate that error companies are more likely to experience liquidity problems. The coefficient of the variable *Big4* is negative and significant (p-value < 0.01), which indicates an inverse relationship between auditor size and the presence of enforcement findings, suggesting that Big 4 auditors might deliver higher audit quality.

There is no significant association between *ClientImp* and *Error*, leading to the rejection of H3 that client importance is negatively associated with audit quality²⁵¹. The obtained results do not show quality issues when the auditor obtains more than 15% of his revenue from the audited entity, so that no evidence can be produced in favour or against the 15% threshold set by art. 4 para.3 EU-reg. However, as there exist theoretical considerations as to why client importance might impair auditor independence, it might be wise to limit the financial dependence of auditors on a client. There remains the question at which value the limit should be set.

This analysis suffers from three issues²⁵². First, the data differ in a few respects from the requirements of the EU-regulation. Due to these differences it is possible that the already few observations exceeding the 15% threshold are overestimated. Second, the provision of the EU-regulation applies only when the 15% threshold is exceeded for more than three years. However, to avoid further restricting the research sample, I only analyse the consequences of an auditor receiving more than 15% of his revenue from a client in one year. Third, by design, companies can be included multiple times in the sample. It follows that the underlying variables may not be independent. Fourth, small sample size issues and low occurrence rates might also affect the analysis. In the sample only 9.7% (11 obs.) of the companies in the error group and 3.5% (4 obs.) of the companies in the control group contribute to more than 15% of the auditor's revenue²⁵³.

²⁵¹ Conversely, Fiallo/Hecker (2019b), pp. 295-296, find weak evidence that client importance is associated with the existence of enforcement findings. Their univariate conditional logistic regression shows that companies with a fee-to-revenue ratio higher than 15% are more likely of having enforcement findings than companies with fees below the 15% threshold. They interpret this result as a preliminary indication that setting a limit on the revenue an auditor can obtain from a client might be useful to improve audit quality.

²⁵² For a discussion of the limitations see Fiallo/Hecker (2019b), pp. 294-296.

²⁵³ As emphasised by Hosmer/Lemeshow/Sturdivant (2013), p. 246, the explanatory power of the conditional model comes from discordant pairs. In this analysis, there are only 13 discordant pairs.

Similarly, Zimmermann (2008) points out the difficulties in determining whether the 15% limit is appropriate or not, due to the rarity of this occurrence²⁵⁴.

Table 3.13: Multivariate results for H3

	expected sign	coeff. (p-value)	odds ratio
<i>ClientImp</i>	+	0.641 (0.197)	1.899
<i>RoA</i>	-	0.594 (0.609)	1.811
<i>Loss</i>	+	0.418 (0.177)	1.519
<i>Lev</i>	+	1.079* (0.076)	2.942*
<i>CFO</i>	-	-3.333** (0.022)	0.036**
<i>Growth</i>	+	0.023 (0.449)	1.024
<i>Big4</i>	-	-0.895*** (0.007)	0.409***
<i>Change</i>	+	0.150 (0.385)	1.162
<i>IFRSfirst</i>	+	-0.048 (0.523)	0.954
<i>Size</i>	-/+	0.650 (0.354)	1.916
Obs.		226	
Prob > X ²		0.006	
Pseudo R ²		0.159	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff.}$. The analysis is based on a restricted sample of 226 observations.

²⁵⁴ Zimmermann (2008), p. 219. A similar observation is made by Quick/Sattler (2011a), pp. 84-85.

3.5.3 Sensitivity tests

3.5.3.1 Financial companies

As mentioned above, one fourth of the sample (68 obs.) is made up of banks, insurance and real estate companies. I choose to include them in the sample to avoid losing too many valid observations. However, fundamentals of financial companies might differ from those of companies belonging to other sectors. Also, the existence of dissimilarities in the provision of audit and non-audit services (e.g. due to a different regulation) might signify that fees are not directly comparable²⁵⁵. By conducting a logistic regression conditional on matched-pairs, observations are compared within the respective industries, which means that sector differences should not confound the results. However, as a robustness check, all financial companies are now dropped from the sample, while the regressions for the testing of H1, H2 and H3 are reran.

Tables 3.14 and 3.15 present the results for the model M1-M4 employed for testing H1. Only the coefficients and odds ratios of the main variables are reported²⁵⁶. As in the model based on the full sample, there is no significant association between other assurance services and audit quality. Tax services are positively associated with audit quality, although the coefficient of *NASratio_c* is now only significant at the 10% level. Regarding other non-audit services, I obtain similar results: the coefficient of *NAScat_d* is negatively associated with enforcement findings in both models M1 and M4 (p-value < 0.01), while the coefficient of *NASratio_d* is still not significant. While in models M1 through M4 for the full sample, *AuditFees* are significant and positively associated with the existence of enforcement findings, here there are only limited indications of such an association²⁵⁷.

The results for H2 and H3 (untabulated) are similar to those obtained for the full sample, not providing any indication that exceeding the cap or obtaining more than 15% of revenue from a client is associated with audit quality. I conclude that the evidence obtained from the main model regarding H1, H2 and H3 is mostly robust when excluding financial companies. However, the association between the fees paid for audit services and the existence of enforcement findings does not seem to hold.

²⁵⁵ E.g. Lehrbass/Scheipers (2017), p. 1439; Gros (2016), p. 230.

²⁵⁶ The results for the control variables (untabulated) are consistent with those obtained so far.

²⁵⁷ The coefficient of *AuditFees* is significant at the 5% level only in M1 and M3 and only when the provision of NAS is measured with the natural log of fees.

Table 3.14: Multivariate results for H1 - without financial companies – part 1

	expected sign	(M1)	(M1)	(M2)	(M2)	(M3)	(M3)	(M4)	(M4)
		coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio
NAScat_b	-/+	-0.031 (0.475)	0.970	-0.027 (0.466)	0.973	-	-	-	-
NAScat_c	-/+	0.117*** (0.008)	0.890***			0.108*** (0.006)	0.898***		
NAScat_d	-/+	0.135*** (0.006)	0.874***					0.146*** (0.004)	0.864***
AuditFees	-/+	1.222** (0.015)	3.395**	0.469 (0.215)	1.598	0.900** (0.041)	2.460**	0.691 (0.103)	1.995
Obs.		214		214		214		214	
Prob > X ²		0.000		0.000		0.000		0.000	
Pseudo R ²		0.460		0.338		0.394		0.403	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff.}$. The analysis is based on a restricted sample of 214 observations.

Table 3.15: Multivariate results for H1 - without financial companies – part 2

	expected sign	(M1)	(M1)	(M2)	(M2)	(M3)	(M3)	(M4)	(M4)
		coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio
NASratio_b	-/+	1.505 (0.383)	4.502	1.988 (0.240)	7.300				
NASratio_c	-/+	-2.993* (0.070)	0.050*			-3.073* (0.058)	0.046*		
NASratio_d	-/+	-1.266 (0.448)	0.282					-0.769 (0.635)	0.464
AuditFees	-/+	0.559 (0.174)	1.750	0.450 (0.233)	1.568	0.576 (0.150)	1.780	0.410 (0.282)	1.506
Obs.		214		214		214		214	
Prob > X ²		0.000		0.000		0.000		0.000	
Pseudo R ²		0.370		0.344		0.360		0.336	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff.}$. The analysis is based on a restricted sample of 214 observations.

3.5.3.2 Endogeneity

Prior studies caution that audit fees, NAS fees and accounting quality are endogenous variables, due to common factors simultaneously affecting the company's financial reporting and its choice of acquired services²⁵⁸. Markelevich/Rosner (2013) argue that NAS fees, audit fees and fraud "are all associated with audit risk factors"²⁵⁹. The presence of endogeneity can produce biased and inconsistent estimators, which is why prior research highlights the importance of using proxies for unobservable variables²⁶⁰. By identifying confounding factors and including them as control variables, it is possible to reduce the problem of endogeneity. However, there is no guarantee that the identified proxies capture every unobservable effect²⁶¹. I follow here the approach developed by Frank (2000) and calculate the impact threshold for a confounding variable (ITCV) to express how large the correlation between a confounding variable and the predictor, as well as between the confounding variable and the outcome must be, to render the statistically significant coefficient of a predictor not significant²⁶².

Table 3.16 shows in each column the ITCV for the statistically significant fee variables from M1 used for testing H1. The coefficients of *NAScat_c*, *NAScat_d* and *NASratio_c* are quite robust, as an omitted variable should have an impact on the dependent variable two or three times larger than the impact of the strongest variable in the model. Conversely, for *AuditFees* it would only take an omitted variable with 57% (in the model with *NAScat_b*, *NAScat_c* and *NAScat_d*) and 9% (in the model with *NASratio_b*, *NASratio_c* and *NASratio_d*) of the impact that *Size* has, to overturn the results. This shows again that *AuditFees* and *Size* are strongly correlated, but also reveals that the inference might be invalidated by an omitted variable. While the findings regarding *NAScat_c*,

²⁵⁸ Antle/Gordon/Narayanamoorthy/Zhou (2006), pp. 236-237.

²⁵⁹ Markelevich/Rosner (2013), pp. 1612-1613. Consider the following example based on a similar one reported by DeFond/Raghunandan/Subramanyam (2002), p. 1268: poor financial performance might increase the likelihood of misstatements, as managers would try to hide the true economic situation of the company. Concurrently, auditors might charge higher audit fees due to the higher risk related to poor financial performance and the consequent greater audit effort. Also, financial distress might influence the decision to buy NAS: managers might either reduce the amount of NAS to save money or increase the extent of consulting services in an attempt to improve profitability.

²⁶⁰ Wooldridge (2016), p. 462.

²⁶¹ DeFond/Raghunandan/Subramanyam (2002), p. 1268; Frank (2000), p. 149.

²⁶² Frank (2000), pp. 150-156; Larcker/Rusticus (2010), pp. 202-203; Blaylock/Gaertner/Shevlin (2017), p. 168. The ITCV is originally employed in connection with an OLS regression. Recent studies argue that when the two categories that the dependent variable can assume have a proportion ranging between 0.25 and 0.75 (in this case 0.5), the results of the logit model and of the OLS model applied to the binary variable are consistent, which is why the ITCV can be used in combination with dichotomous dependent variables, see Cheng/Martin/Werum (2007a), p. 54. Cheng/Martin/Werum (2007b), pp. 10-14; Hubbard/Christensten/Graffin (2017), p. 2262.

NAScat_d and *NASratio_c* are quite robust, there are indications that the association between audit fees and enforcement findings might be weak.

Table 3.16: ITCV for M1 (*NAScat_c*, *NAScat_d*, *NASratio_c*, *AuditFees*)

	<i>NAScat_c</i>	<i>NAScat_d</i>	<i>NASratio_c</i>	<i>AuditFees</i> (1)	<i>AuditFees</i> (2)
<i>AuditFees</i>	0.0205	0.0033	0.0072		
<i>Big4</i>	0.0031	-0.0167	-0.0044	-0.0258	-0.0293
<i>CFO</i>	0.0080	0.011	0.0116	-0.0218	-0.0175
<i>Change</i>	-0.0035	-0.006	-0.0073	-0.0019	0.0054
<i>Growth</i>	0.0010	0.0043	-0.0048	-0.0037	0.0012
<i>IFRSfirst</i>	0.0008	0.0011	0.0001	0.0015	0.0012
<i>Lev</i>	-0.0060	0.0042	-0.0021	0.0256	0.0203
<i>Loss</i>	-0.0050	0.0012	-0.0066	0.0076	
<i>NAScat_b</i>	0.0004	0.0006		0.0075	
<i>NAScat_c</i>		-0.0373		-0.0313	
<i>NAScat_d</i>	-0.0321			-0.0036	
<i>NASratio_b</i>			-0.0116		0.0029
<i>NASratio_c</i>					-0.0109
<i>NASratio_d</i>			0.0006		0.0057
<i>RoA</i>	-0.0003	-0.0008	-0.0017	0	-0.002
<i>Size</i>	0.0007	-0.01	0.003	0.1311	0.0807
Largest impact	-0.0321	-0.0373	-0.0116	0.1311	0.0807
Largest impact variable	<i>NAScat_d</i>	<i>NAScat_c</i>	<i>NASratio_b</i>	<i>Size</i>	<i>Size</i>
ITCV	-0.1017	-0.0812	-0.0392	0.0742	0.0072
Magnitude	3.17	2.18	3.38	0.57	0.09

Notes: The table shows the ITCV index for the fee variables with significant coefficients according to M1 (*NAScat_c*, *NAScat_d*, *NASratio_c*, *AuditFees*). The ITCV is the minimum impact an omitted variable should have to render the coefficient of the variable of interest insignificant. It is measured as the product of the partial correlation between the variable of interest and the confounding variable and the partial correlation between the dependent variable and the confounding variable. Magnitude expresses the size of the impact compared to the largest impact variable. For the variable *AuditFees* the ITCV is estimated alternatively for M1 using *NAScat_b*, *NAScat_c*, *NAScat_d* (1) and *NASratio_b*, *NASratio_c*, *NASratio_d* (2).

3.5.3.3 Size

In each model, the company's size is accounted for with the variable *Size*, which is measured as the natural log of total assets. I discussed before the importance of including the matching variable when the matching is "imperfect", like in the case where pairs are of a similar, but not the same size²⁶³. Even though it may be more appropriate to include the exact matching variable in the model, in this case total assets, I have chosen to employ the natural log of total assets, since I expect

²⁶³ See the considerations made in section 3.4.4.

size to behave non-linearly in the model²⁶⁴. There are two caveats with this choice: one, the imperfect matching is only partially accounted for and two, *AuditFees* and *Size*, which are respectively the natural logarithm of audit fees and of total assets, are strongly positively correlated ($r = 0.89$)²⁶⁵. I rerun all regressions for the testing of H1, H2 and H3 measuring size with total assets²⁶⁶. The correlation between *AuditFees* and total assets is still positive, but significantly less strong ($r = 0.366$). Additionally, using total assets allows to account for all the residual pair-wise difference in size. The results (untabulated) remain consistent with those obtained so far, but the explanatory power of the models increases considerably. I conclude that the results are not biased by the choice of using the natural logarithm of total assets in place of total assets as a control variable.

3.6 Conclusion

The recently enacted EU-regulation on the statutory audit of PIEs aims at improving audit quality by limiting the provision of certain NAS (black list) and by setting a cap on NAS fees as well as a limit on the proportion of fees an auditor can obtain from a client. Whether or not auditors should provide certain NAS to their clients is a much-discussed topic. From a theoretical point of view, the provision of NAS might strengthen the economic bond as it increases the level of quasi-rents an auditor obtains from a client. At the same time, the additional knowledge generated by NAS might boost the ability of the auditor to detect misstatements. It follows that, at least in theory, the provision of NAS might both improve and reduce audit quality, which raises the question of how the matter should be regulated.

In this paper, I have addressed three questions related to the provision of NAS, and the level of fees charged by the auditor. First, I questioned if and how certain types of services affect audit quality. While there is no evidence that other assurance services are associated with audit quality, there are indications that the provision of other non-audit services and of tax services might improve audit quality thanks to knowledge spillovers. The analysis also shows that audit fees are negatively associated with audit quality, although the sensitivity tests suggest that these results might not be robust. These findings indicate that, even though there might be a risk of economic bonding when

²⁶⁴ On the trade-off between fully accounting for the imperfect matching and modelling the variable's contribution see Cram/Karan/Stuart (2009), pp. 489, 503.

²⁶⁵ However, the VIFs do not reveal any multicollinearity issues, see section 3.5.1.

²⁶⁶ A similar robustness test is done e.g. in Fiallo/Hecker (2019a), p. 233 footnote no. 77.

the auditor receives high fees from a client, this risk is unlikely to come from tax and other non-audit services. On the contrary, the results show that tax services, and possibly also other non-audit services, might have the potential of increasing audit quality by providing additional client-specific knowledge. For this reason, I do not identify the need of a stricter regulation on NAS in Germany, as in the period prior to the EU-regulation the analysis does not show any audit quality issue related to the provision of NAS²⁶⁷. Extending the prohibitions on NAS might reduce the occurrence of knowledge spillovers with a consequent decrease in audit quality²⁶⁸. In this sense, the decision of the German regulator to make the black list less extensive by allowing some tax services seems to be appropriate²⁶⁹.

Second, I analysed whether there is an association between violating the cap on NAS fees and audit quality. The results do not show that exceeding the cap is associated with the existence of enforcement findings, thus not providing any evidence on whether the introduction of the cap is a useful measure to improve auditor independence and audit quality. Prior studies question the ability of the cap to have a significant impact on audited companies²⁷⁰. Similarly, this study shows that in 84.7% of the analysed cases, companies had fees below the cap or did not qualify for its application²⁷¹. As the analysis in this study concerns the period preceding the implementation of the black list, the potential impact of the cap might be even smaller once the provisions of art. 5 EU-reg. apply²⁷². Given its potentially limited application as well as the lack of evidence supporting the choice of the 70% threshold, the question of if and to what extent audit quality might benefit from the introduction of the cap remains unanswered.

Third, I addressed the question of whether audit quality is lower when the auditor obtains more than 15% of his revenues from a client. I do not find any evidence that the auditor is more likely to impair his independence for economically important clients²⁷³, while I observe that due to the

²⁶⁷ A similar interpretation of the results is provided by Fiallo/Hecker (2019a), p. 234.

²⁶⁸ A similar concern is expressed by Ratzinger-Sakel/Schönberger (2015), pp. 62-63.

²⁶⁹ A similar interpretation of the results is provided by Fiallo/Hecker (2019a), p. 234 and by Hohenfels/Quick (2018), OnlineFirst 25 October 2018. Available at: <https://link.springer.com/article/10.1007/s11846-018-0306-z>.

²⁷⁰ E.g. Fiallo/Hecker (2019c), p. 347; Ratzinger-Sakel/Schönberger (2015), p. 62; Pott/Schröder/Weckelmann (2014), p. 502. Other researchers also suggest that the 70% threshold might be too high, see e.g. Hohenfels/Quick (2018), OnlineFirst 25 October 2018. Available at: <https://link.springer.com/article/10.1007/s11846-018-0306-z>.

²⁷¹ 89% in the error group and 81% in the control group, see Table 3.3. See also the considerations made on the potential impact of the cap in Fiallo/Hecker (2019b), p. 297.

²⁷² Fiallo/Hecker (2019b), pp. 292-293, make a similar consideration.

²⁷³ Conversely, Fiallo/Hecker (2019b), pp. 295-296, find weak evidence that client importance is associated with the existence of enforcement findings. Their univariate conditional logistic regression shows that companies with a fee-

existing regulation in Germany, it is rather rare for an auditor to obtain more than 15% of his revenue from a client²⁷⁴. Under these conditions, it might be difficult to provide evidence in favour or against the 15% threshold set by art. 4 para. 3 EU-reg. However, as there are theoretical considerations as to the why client importance might impair auditor independence, it might be wise to limit the financial dependence of an auditor on a client. There remains the question at which value the limit should be set²⁷⁵.

This study suffers from some limitations²⁷⁶. First, as audit quality cannot be directly measured, enforcement announcements are employed as a proxy. The conclusions drawn from this analysis depend on the ability of the surrogate to correctly reflect the actual audit quality. Second, the capacity to infer association between the level of fees and the existence of enforcement findings rests on the reliability of the underlying data. Prior studies point out the presence of a bias due to the inaccurate disclosure of fees by the audited entities²⁷⁷. Additionally, ambiguities in the disclosure, as well as the lack of official guidance on the classification into categories have led to inconsistencies in practice²⁷⁸. These problems have been aggravated by changes to the regulation over time such as the introduction of BilMoG. To minimise the fee bias, data from the annual reports are hand collected and the pre and post BilMoG era is accounted for in some parts of the analysis²⁷⁹. A further source of measurement errors is due to the discrepancies between the requirements of the EU-regulation and the available data. Third, this study is set in Germany before the introduction of the EU-regulation, which is why the results might not apply to other countries and should be interpreted considering the existing national regulation between 2005 and 2017. Finally, the results are strongly dependent on the research design (matched pairs) as well as on the

to-revenue ratio higher than 15% are more likely of having enforcement findings than companies with fees below the 15% threshold. They interpret this result as a preliminary indication that setting a limit on the revenue an auditor can obtain from a client might be useful to improve audit quality.

²⁷⁴ On the potential impact of the regulatory measures, see the criticism raised by Fiallo/Hecker (2019b), p. 297.

²⁷⁵ See here also the issue pointed out by Fiallo/Hecker (2019c), pp. 347-348, on how the intent of the regulator can be bypassed when services are provided by the auditor's affiliated companies.

²⁷⁶ For a discussion of the limitations, see Fiallo/Hecker (2019c), p. 348.

²⁷⁷ E.g. Markelevich/Rosner (2013), p. 1621; Fiallo/Hecker (2019a), p. 231.

²⁷⁸ E.g. Köhler/Ratzinger-Sakel (2012), p. 286; Quick/Sattler (2011a), p. 78; Sattler (2011) pp. 325-326; Krauß/Zülch (2013), p. 322.

²⁷⁹ The possible confounding effect due to the changes in the interpretation of the national disclosure requirements following the introduction of BilMoG has been accounted for by adding an interaction term in the regression models, to account for the period before and after BilMoG.

choice of control companies. Future studies might try different research designs and different matching approaches²⁸⁰.

²⁸⁰ See also a similar suggestion for a sensitivity analysis of Caliendo/Kopeinig (2008), pp. 44-45.

4. Future fees, current NAS fees and audit quality: Evidence based on enforcement announcements in Germany

Abstract: According to *DeAngelo* (1981a), the auditor-client dynamic should be analysed in a multi-period framework, as it is the expected present value of quasi-rents from future contracts (economic bond) which affects current auditor's decisions. Conversely, the current provision of NAS should not affect auditor independence but might generate knowledge spillovers, thus increasing the ability of the auditor to detect errors. In this paper, I examine how the current provision of NAS as well as the expectations regarding the future provision of services affect audit quality. The results show that future fees are negatively associated with audit quality, suggesting that the expected provision of services might impair auditor independence. Interestingly, the economic bond seems to be driven by the level of future audit fees. Current NAS fees have a positive effect on audit quality, which is consistent with the existence of knowledge spillovers. Considering this evidence, there seems to be no need for further limiting the provision of NAS, as this could reduce audit quality. Also, as there are only limited indications that future NAS fees are negatively associated with audit quality, it is unclear if further restricting the provision of NAS or putting a cap on NAS fees might help increasing auditor independence.

I appreciate the helpful suggestions and comments of Renate Hecker who motivated me for this paper. It builds on part of the descriptive and univariate analyses in Fiallo/Hecker (2019a, 2019b and 2019c) and benefits from part of the data and analyses, and from the thoughts of these three former publications as well as from the fruitful discussions we had while writing these publications.

4.1 Introduction

Recently, the European Union has expressed concerns regarding the provision of non-audit services (NAS)¹ and the level and structure of the fees paid by audited companies to their statutory auditor². Audit and non-audit fees might create an economic bond between auditor and client, which might compromise auditor independence and negatively affect audit quality. The risk of economic bonding has been crystallised in the accounting literature by the quasi-rent theory. According to *DeAngelo* (1981a), while the initial audit might be characterised by fees below the actual cost of audit (low balling), subsequent audits generate positive quasi-rents, due to the cost advantage the incumbent auditor has, compared to other auditors³. The incumbent auditor therefore has an economic interest in keeping the client, which might threaten his independence. The expected provision of NAS, on the one hand, might increase the quasi-rents, as the auditor has an additional cost advantage due to the NAS related start-up and switching costs, thus strengthening the economic bond⁴. On the other hand, the provision of additional services in the current period allows the auditor to gain client-specific knowledge that might improve his ability to detect errors (knowledge spillovers)⁵.

In this paper, I examine how the current provision of NAS as well as the expectations regarding the future provision of audit and non-audit services affect audit quality. According to *DeAngelo* (1981a) the auditor-client dynamic should be analysed in a multi-period framework, as it is the expected present value of quasi-rents from future contracts which affects the auditor's decisions⁶. Prior studies mostly focus only on the current period, leaving a gap between theory and applied research⁷.

¹ Recital 8 Regulation (EU) No 537/2014. For a detailed analysis of certain aspects of Regulation (EU) No 537/2014, including the comparison with the (prior) national regulation in Germany, see Fiallo/Hecker (2019a), p. 230 and Fiallo/Hecker (2019b), pp. 290-291, 294.

² Recital 7 Regulation (EU) No 537/2014.

³ *DeAngelo* (1981a), pp. 113-127. This theory, as well as its extensions, are discussed in more details in section 3.3.1.2.

⁴ *Ostrowski/Söder* (1999), p. 562; *DeAngelo* (1981a), p. 113. On the interaction between auditor independence (or lack thereof) and audit quality see also Fiallo/Hecker (2019a), p. 226.

⁵ *Simunic* (1984), pp. 680-681. On the potential benefits of knowledge spillovers on audit quality, see e.g. Fiallo/Hecker (2019a), p. 231. A more detailed analysis of the theoretical arguments beyond knowledge spillovers is included in section 3.3.1.2 of this study.

⁶ *DeAngelo* (1981a), p. 115.

⁷ A similar consideration is made by *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018), p. 4, and *Blay/Geiger* (2013), p. 583. Besides their research, there are a few other studies which analyse the association between future fees or fee growth and auditor independence, e.g.: *Fuhrmann* (2016), *Causholli/Chambers/Payne* (2014), *Sattler* (2011), *Cahan/Emanuel/Hay/Wong* (2008). Their findings are presented in section 4.2.2.

This study contributes to the existing literature on auditor's fees and audit quality by researching the relation between future fees and current auditor's decisions. Following *Blay/Geiger* (2013), *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018) and *Fuhrmann* (2016), I also examine the provision of services in a multi-period framework, as I argue this might help to separate the effect of the economic bond from that of knowledge spillovers. By using this multi-period approach, I try to provide some answers to the concerns expressed by the European regulator regarding the provision of certain NAS and the level and structure of the fees paid by audited companies to their statutory auditor⁸.

The results show that future fees are negatively associated with audit quality, suggesting that the expected provision of audit services and NAS might impair auditor independence and lead to lower audit quality. This finding is in line with that of *Blay/Geiger* (2013) and is consistent with the existence of economic bonding⁹. Interestingly, the economic bond seems to be driven by the level of future audit fees, while the results provide limited support that future NAS fees are negatively associated with audit quality¹⁰. Regarding the current provision of NAS, contrary to what was reported by *Blay/Geiger* (2013)¹¹ and *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018)¹², I find evidence that current NAS fees have a positive effect on audit quality, which is consistent with the provision of NAS in the current period generating knowledge spillovers. This is in line with the findings discussed in section 3.5.2.1 as well as in *Fiallo/Hecker* (2019a)¹³, where there are indications that certain NAS, e.g. tax services, might improve audit quality. Considering this evidence, there seems to be no need for further limiting the provision of NAS, as this could reduce audit quality. Also, as there are only limited indications that future NAS fees are associated with enforcement findings, it is unclear if reducing the possibility to provide NAS or putting a cap on NAS fees might help increasing auditor independence.

⁸ These concerns as well as certain aspects of the Regulation (EU) No 537/2014 are discussed in section 3.2 as well as in *Fiallo/Hecker* (2019a), p. 230 and *Fiallo/Hecker* (2019b), pp. 290-291, 294.

⁹ *Blay/Geiger* (2013), p. 592.

¹⁰ These results are only partially in line with those of *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018), pp. 22-24, who find that audit quality is likely to be affected by the expectations of future purchase of other non-audit services, but not by the future provision of tax and audit-related services. Conversely, *Fuhrmann* (2016), p. 57, does not find any significant association between future NAS fees and audit quality.

¹¹ *Blay/Geiger* (2013), p. 598.

¹² *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018), pp. 22-24.

¹³ *Fiallo/Hecker* (2019a), pp. 232-234.

This study complements the findings of *Blay/Geiger* (2013), on financially distressed US manufacturing companies¹⁴, *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018) on Spanish listed companies¹⁵ and *Fuhrmann* (2016) on German companies listed on the CDAX¹⁶, by using a sample of German companies subject to enforcement findings between July 2005 and December 2017, and a control group¹⁷. Also, this study provides additional evidence on the relation between the current and expected provision of services and audit quality, by using another proxy for audit quality, namely enforcement announcements¹⁸. Audit quality is measured with the existence of accounting errors in audited financial statements which received an unmodified opinion. Accounting errors are identified thanks to the enforcement announcements, which report any material errors found by the enforcement institutions during their examinations¹⁹.

This paper is organised as follows: section 4.2 outlines the theoretical background, the development of the hypotheses and describe prior research. Section 4.3 presents the methodology of the research, whose results are discussed in section 4.4. I conclude in section 4.5.

4.2 Theoretical background, research hypotheses and prior research

4.2.1 Theoretical background and hypotheses

According to the definition of *DeAngelo* (1981a), audit quality is the auditor's perceived ability to detect misstatements and to report them²⁰. The ability to detect errors depends on the auditor's competence as well as on the correct execution of the necessary audit procedures. The willingness of the auditor to report the errors is conditional on his independence. *DeAngelo* (1981a) argues that the incumbent auditor is able to earn quasi-rents from a client, as he has a cost advantage compared

¹⁴ *Blay/Geiger* (2013), pp. 580-581.

¹⁵ *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018), pp. 4-6.

¹⁶ *Fuhrmann* (2016), p. 57. Other studies focus on NAS growth: *Cahan/Emanuel/Hay/Wong* (2008); *Causholli/Chambers/Payne* (2014); *Sattler* (2011).

¹⁷ This is a reduced sample of that used in *Fiallo/Hecker* (2019a), p. 228, and in chapter 3.

¹⁸ The studies of *Blay/Geiger* (2013), *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018) and *Fuhrmann* (2016) have employed e.g. going concern opinions, modified opinions, restatements, earnings surprises, meeting or beating earnings benchmark or analyst's forecasts and the propensity to report a small increase in profit compared to the previous year.

¹⁹ For a discussion of the arguments in favour and against the use of this surrogate see sec. 3.4.1 as well as *Fiallo/Hecker* (2019a), p. 226. There are to my knowledge a few other studies on NAS fees, which use enforcement announcements as a surrogate for audit quality: e.g. that of *Markelevich/Rosner* (2013) set in the US, and that of *Gros* (2016) and *Krauß/Pronobis/Zülch* (2015) set in Germany. However, they do not employ a multi-period framework.

²⁰ *DeAngelo* (1981b), p. 186. This definition provided by *DeAngelo* is commonly used in auditing literature, see e.g. *Fiallo/Hecker* (2019a), p. 226. The theory of *DeAngelo* and its extensions are discussed in section 3.3.1.

to other potential auditors²¹. Under these circumstances, independence is threatened by the expected present value of quasi-rents deriving from future contracts. The model of *DeAngelo* (1981a) outlines the auditor-client dynamic in a multi-period framework, where the current auditor's decisions are affected by the expectations of future quasi-rents²². Nonetheless, prior studies mainly analyse how fees paid for audit and non-audit services in the current period affect audit quality, leaving a gap between theory and applied research²³.

Figure 4.1 illustrates the relation between audit quality and the provision of NAS in a multi-period framework. In the current period, the provision of NAS might generate knowledge spillovers, which might improve the auditor's ability to detect errors²⁴. His willingness to report them, however, would not be affected by the current provision of NAS but by the quasi-rents the auditor expects to obtain from future contracts. By providing additional NAS, the value of the quasi-rents might increase, also due to cost savings related to the joint provision of audit and non-audit services²⁵, also increasing the risk of economic bonding. It follows that, examining the provision of NAS in a multi-period framework might help to separate the effect of the economic bond from that of knowledge spillovers²⁶.

²¹ DeAngelo (1981a), pp. 118-123. The cost advantage is due to the higher initial costs a new auditor would face and to the transaction costs, the audited company would sustain when switching auditor. For more details on the assumptions and consequences of the existence of quasi-rents, see section 3.3.1.2.

²² DeAngelo (1981a), pp. 118-123.

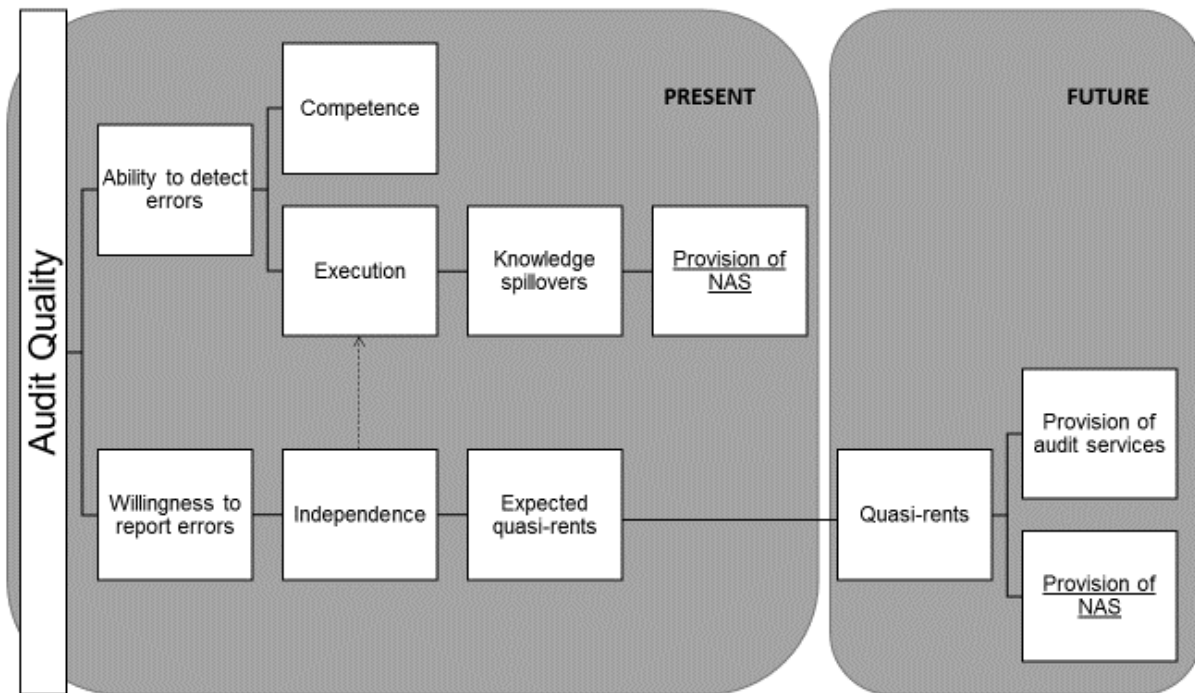
²³ A similar consideration is made by Castillo-Merino/Garcia-Blandon/Martinez-Blasco (2018), p. 4, and Blay/Geiger (2013), p. 583. Besides their research, there are a few other studies which analyse the association between future fees or fee growth and auditor independence, e.g.: Causholli/Chambers/Payne (2014), Fuhrmann (2016), Sattler (2011), Cahan/Emanuel/Hay/Wong (2008). Their findings are presented in section 4.2.2.

²⁴ On the potential benefits of knowledge spillovers on audit quality see Fiallo/Hecker (2019a), p. 231. A more detailed analysis of the theoretical arguments behind knowledge spillovers is included in section 3.3.1.2.

²⁵ Beck/Frecka/Solomon (1988), pp. 57-61; Ostrowski/Söder (1999), pp. 559-560. As in the market for audit services, the incumbent auditor has a cost advantage due to NAS related start-up and switching costs.

²⁶ It may not be possible to entirely isolate them. E.g. DeAngelo (1981a), p. 116, points out that in case of impaired independence, the auditor would probably adjust his audit effort accordingly, having no interest in finding any error. This issue is also mentioned in Fiallo/Hecker (2019a), pp. 226 and 231.

Figure 4.1: Audit quality and the provision of NAS in a multi-period framework



In this study, I examine how the current provision of NAS as well as the expectations regarding the future provision of audit services and NAS affect audit quality. Audit quality is proxied by the existence of enforcement findings in financial statements which received an unmodified opinion²⁷. The expected value of quasi-rents is approximated by future fees. Here two assumptions are made: first, that future fees are positively correlated with quasi-rents²⁸. The use of a proxy is necessary because quasi-rents are not directly observable, as besides the level of fees, they also depend on the client specific costs and their development in the future²⁹; second, that future fees correctly approximate the auditor's current expectations³⁰. Finally, to capture knowledge spillovers, I use current NAS fees, assuming that the higher the fees, the more extensive are the services provided by the auditor and in turn the higher the probability to generate knowledge spillovers.

H1: There is a positive association between enforcement announcements and future fees.

H2: There is a negative association between enforcement announcements and current NAS fees.

²⁷ For a discussion on the difficulties in measuring audit quality as well as the limitations to the use of this proxy, see sec. 3.4.1 as well as Fiallo/Hecker (2019a), p. 226.

²⁸ This assumption is also reported by Wagenhofer/Ewert (2015), p. 525.

²⁹ Wagenhofer/Ewert (2015), p. 525.

³⁰ A similar assumption is made by Blay/Geiger (2013), p. 583.

4.2.2 Prior literature

There are to my knowledge only a few studies, which have analysed the relation between future fees and audit quality³¹. The study of *Blay/Geiger* (2013), on financially distressed US manufacturing companies over the period 2004 to 2006, finds that auditor independence in the current period is impaired by the expectation of future benefits³². They collect evidence that future fees, measured as the natural logarithm of the fees paid in the two subsequent periods, are negatively associated with the issue of a going concern opinion, which seems to indicate that auditors are more favourable towards clients paying higher future fees. Also, they find that current NAS fees are negatively associated with the issue of a going concern opinion, which they interpret as a sign of economic bonding.

Fuhrmann (2016) investigates the relation between future NAS fees and audit quality, measured as the propensity of the firm to meet or beat analyst's forecasts, as well as to report a small increase in profit compared to the previous year. Her research sample includes all German companies listed as of December 2015 on the CDAX, for which she collects data for the years 2007-2015. The results provide no evidence of an association between future NAS fees and audit quality³³.

Castillo-Merino/Garcia-Blandon/Martinez-Blasco (2018) research the effect the current and expected provision of NAS have on audit quality, measured with restatements and earnings surprises³⁴. Their results, for a sample of Spanish listed companies between 2005 and 2016, show evidence that audit quality is affected by the expectation of future NAS purchases. Specifically, it is future fees for other non-audit services which are positively associated with restatements and the propensity to meet or beat analysts' earnings forecasts, while there is no significant association with future tax and audit-related fees³⁵. They find no significant relation between restatements and current NAS fees and only weak evidence that current NAS fees are negatively associated with the

³¹ For a review of the literature on the provision of NAS, see section 3.3.2.

³² *Blay/Geiger* (2013), pp. 580-581.

³³ *Fuhrmann* (2016), p. 57.

³⁴ *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018), pp. 4-6. In their sensitivity analysis, they also use meeting or beating earnings benchmark and audit opinions as proxy for audit quality, see *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018), pp. 25-29.

³⁵ *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018), pp. 22-24.

propensity to meet or beat analysts' earnings forecasts. Conversely, their results show that audit quality is negatively associated with current audit fees³⁶.

Other studies focus on NAS growth. *Cahan/Emanuel/Hay/Wong* (2008) investigate whether NAS growth, measured as the average annual growth rate for NAS fees in the following two, four and five years, is associated with earnings management, measured with discretionary accruals³⁷. Their analysis, based on a sample of firms listed on the New Zealand Exchange during the period 1995-2001, does not show any significant association.

Causholli/Chambers/Payne (2014) analyse the association between earnings management and the opportunity for NAS fee growth for a sample of US companies during the pre-SOX period 2000-2001³⁸. Their analysis is done at partner level, as they expect independence issues to be affected by partner short-term profitability goals rather than firm reputation concerns. They discover that clients, with relatively low NAS in the current period and an increase in the purchase of NAS in the following period, have larger discretionary accruals in the current period. This might indicate that the opportunity for selling additional NAS in the future impairs current auditor independence. Additionally, they find that the negative effect of future NAS on earnings management is stronger for companies with incentives to manage earnings and weaker for companies with stronger corporate governance. Regarding the fees paid in the current period, they observe that NAS fees are negatively associated with income-increasing accruals, which is consistent with NAS creating knowledge spillovers. However, their results cannot be replicated when using restatements and going concern opinions as alternative measures of audit quality.

Finally, *Sattler* (2011) analyses whether the change in NAS, measured as a percentage change in the fees paid for NAS in t and $t+1$, is associated with discretionary accruals among German companies listed on the DAX, MDAX, SDAX and TecDax for the years 2005 to 2007³⁹. The results do not show any significant association between the change in NAS and the presence of earnings management; the same is true when NAS growth is considered separately according to the service categories (other assurance services, tax services and other non-audit services).

³⁶ Castillo-Merino/Garcia-Blandon/Martinez-Blasco (2018), pp. 22-24.

³⁷ Cahan/Emanuel/Hay/Wong (2008), p. 182.

³⁸ Causholli/Chambers/Payne (2014), pp. 682-683.

³⁹ Sattler (2011), pp. 409-412.

4.3 Research design

4.3.1 Sample selection

Companies subject to enforcement findings are identified based on the enforcement announcements published in the Federal Gazette (Bundesanzeiger)⁴⁰. The research sample includes 131 erroneous financial statements for financial years between 2005 and 2015 (Table 4.1).

Table 4.1: Sample selection

Enforcement announcements issued on the Federal Gazette until 31/12/2017	258
Announcements disclosed twice or amended	- 17
Announcements regarding half-year financial statements	- 30
Announcements regarding foreign companies	- 19
Announcements regarding annual financial statements (local GAAP)	- 11
Announcements regarding consolidated financial statements (US-GAAP or local GAAP)	- 8
Announcements regarding financial statements which received an adverse auditor's opinion	- 3
Announcements regarding financial statements which received a qualified auditor's opinion	- 19
Announcements regarding financial statements beginning before 01/01/2005	- 7
Announcements regarding financial statements for which fee data are not available	- 13
Final sample	131

Notes: This table is based on Fiallo/Hecker (2019a), p. 228 Table 2. The differences in the sample size are explained in footnote no. 40.

In order to analyse the relationship between future fees and enforcement findings, each error company is matched in the error year to a control company⁴¹. The choice of the control group aims at matching together German companies with similar characteristics in terms of size, sector and applicable accounting standards (IFRS). Control companies are also subject to the enforcement regime, but have not disclosed any enforcement announcements until 31 December 2017, which suggests they might have higher accounting and audit quality compared to error companies⁴².

The sector and year distribution in the sample is presented in Table 4.2. One fourth of the sample (66 cases) is made up of financial companies (credit institutions, insurance and real estate companies). As discussed in section 3.4.2, prior accounting studies tend to exclude financial

⁴⁰ For more details on the selection of the sample, see section 3.4.2. The same approach is also used in Fiallo/Hecker (2019a), p. 228. Compared to the sample described in Fiallo/Hecker (2019a), p. 228 Table 2, there are 20 fewer observations: 7 announcements regarding financial years beginning before 1 January 2005 are eliminated, as companies were not required to disclose auditor fees; 13 announcements are eliminated as data on current and future fees were not available. Among them two sets of financial statements for the years 2005 and 2015 are excluded as they report only the fees paid to the worldwide network of the group auditor.

⁴¹ The details on the matching approach are discussed in section 3.4.2 as well as in Fiallo/Hecker (2019a), p. 228.

⁴² For the limitations to this statement, see section 3.4.2 as well as Fiallo/Hecker (2019a), pp. 226-227.

companies from the analysis, as certain accounting values and the fees charged might not be comparable with other sectors⁴³. To make sure that their inclusion does not drive the results, these observations are dropped from the sample in a sensitivity test.

Table 4.2: Sector and year distribution

Sector	No.	Perc.	Year	No.	Perc.
Consumer Discretionary	58	22.1%	2005	34	13%
Financials	66	25.2%	2006	46	17.6%
Telecom Services	8	3.1%	2007	20	7.6%
Materials	6	2.3%	2008	36	13.7%
Information Technology	52	19.8%	2009	36	13.7%
Consumer Staples	6	2.3%	2010	16	6.1%
Health Care	20	7.6%	2011	14	5.3%
Industrials	40	15.3%	2012	26	9.9%
Utilities	6	2.3%	2013	10	3.8%
			2014	14	5.3%
			2015	10	3.8%
Total	262	100%		262	100%

Notes: Companies are classified according to the Global Industry Classification Standard (GICS).

4.3.2 The model

To test how the current provision of NAS as well as the expectations regarding the future provision of audit and non-audit services affect audit quality, I run a conditional logistic regression, which is recommended for matched samples, as it allows to determine the likelihood of an outcome when the observations are not independent but paired⁴⁴. The model is the following⁴⁵:

$$\begin{aligned}
 Error_{i,t} = & \beta_1 * FutureFees_{i,t} + \beta_2 * NASfees_{i,t} + \beta_3 * RoA_{i,t} + \beta_4 * Loss_{i,t} + \beta_5 * Lev_{i,t} \\
 & + \beta_6 * CFO_{i,t} + \beta_7 * Growth_{i,t} + \beta_8 * Big4_{i,t} + \beta_9 * Change_{i,t} \\
 & + \beta_{10} * IFRSfirst_{i,t} + \beta_{11} * Size_{i,t} + \varepsilon_{i,t}
 \end{aligned}$$

$Error^{46}$ is the binary dependent variable, which assumes the value of 1 if the financial statements for the year t of company i are subject to an enforcement announcement, and 0 otherwise. The

⁴³ E.g. Lehrbass/Scheipers (2017), p. 1439; Gros (2016), p. 230.

⁴⁴ Hosmer/Lemeshow/Sturdivant (2013), pp. 243-251; Cram/Karan/Stuart (2009), p. 480; Stuart/Shin/Cram/Karan (2013), pp. 89-90. See also the similar considerations reported by Fiallo/Hecker (2019a), p. 233.

⁴⁵ In choosing the model variables, I keep in mind the rule of thumb that, to achieve a reliable estimation of the coefficients in a logistic regression, the number of covariates should stay at least in a 1:10 proportion to the least frequent outcome, in this case the presence (or lack) of enforcement findings. See Hosmer/Lemeshow/Sturdivant (2013), pp. 407-408. Due to data constraints, in some of the sensitivity tests, the proportion is slightly under the above value.

⁴⁶ Hereafter, apart from Table 4.3 and Table 4.4, the variables are reported without subscripts.

independent variable (*FutureFees*) represents the expected quasi-rents and should encompass the economic interest the auditor has in the client. As in prior studies, I measure the variables *Future_AF*, *Future_NAF* and *Future_TF* as the natural logarithm of the fees paid in the subsequent two years to the statutory auditor for audit services and non-audit services, as well as their total⁴⁷. The decision to use a two-year period is motivated by the comparatively higher evaluation of expected benefits in the near future⁴⁸. Additionally, a two-year window seems appropriate as the auditor's expectations might be closer to the fees actually charged, while they might become less accurate with time. Future audit and non-audit fees are analysed together (M3) as well as separately (M1 and M2)⁴⁹. The variable *NASfees*, which is the natural log of the fees paid for non-audit services in the current period, should capture the existence of knowledge spillovers. *NASfees* is employed in each model (M1-M3). Table 4.3 provides an overview of the fee variables.

Table 4.3: Fee variables

Variables	Description	Effect
<i>Future_AF_{i,t}</i>	Natural log of audit fees disclosed by company i in year t+1 and t+2.	Positive
<i>Future_NAF_{i,t}</i>	Natural log of non-audit fees disclosed by company i in year t+1 and t+2.	Positive
<i>Future_TF_{i,t}</i>	Natural log of total fees disclosed by company i in year t+1 and t+2.	Positive
<i>NASfees_{i,t}</i>	Natural log of non-audit fees disclosed by company i in year t.	Negative

Notes: The expected effect is estimated referring to the probability of having enforcement findings. All data underlying the variables have been manually collected from consolidated financial statements.

The model controls for several factors affecting the probability of error such as the company's economic situation, firm complexity, and several auditor-related characteristics⁵⁰. The control variables are listed in Table 4.4.

⁴⁷ E.g. Blay/Geiger (2013), p. 587; Castillo-Merino/Garcia-Blandon/Martinez-Blasco (2018), p. 9.

⁴⁸ Blay/Geiger (2013), p. 587.

⁴⁹ E.g. Larcker/Richardson (2004), p. 632, suggest analysing them separately as quasi-rents might be higher for non-audit services than for audit fees.

⁵⁰ A more detailed discussion on the control variables is included in section 3.4.4.

Table 4.4: Control variables

Variables	Description	Effect
$RoA_{i,t}$	Ratio of net income to total assets for company i in year t.	Negative
$Loss_{i,t}$	Dichotomous variable coded 1, if company i reports a loss in year t, and zero otherwise.	Positive
$Lev_{i,t}$	Ratio of total liabilities to total assets for company i in year t.	Positive
$CFO_{i,t}$	Ratio of operating cash-flow to total assets for company i in year t.	Negative
$Growth_{i,t}$	Percentage change in revenues for company i in year t compared to t-1.	Positive
$IFRSfirst_{i,t}$	Dichotomous variable coded 1, if company i in year t applied IFRS for the first time, and zero otherwise.	Positive
$Big4_{i,t}$	Dichotomous variable coded 1, if the group auditor of company i in year t is one of the Big 4, and zero otherwise.	Negative
$Change_{i,t}$	Dichotomous variable coded 1, if the group auditor of company i in year t is appointed for the first time, and zero otherwise.	Positive
$Size_{i,t}$	Natural log of total assets for company i in year t.	Positive/ Negative

Notes: The expected effect is estimated referring to the probability of having enforcement findings. Data underlying the following variables have been manually collected from consolidated financial statements: $IFRSfirst_{i,t}$, $Big4_{i,t}$, $Change_{i,t}$. All other underlying data have been retrieved from Thomson Reuters Eikon, where they are available under the following codes: TR.Revenue, TR.TotalLiabilities, TR.TotalAssets, TR.NetIncomeBeforeExtraItems, TR.CashFromOperatingAct.

4.4 Results

4.4.1 Descriptive statistics

When performing univariate tests, it is important to account for paired observations⁵¹. Therefore, I run separated regressions for each variable with dummies for pairs and using the existence of enforcement findings and size as explanatory variables⁵². All continuous variables are winsorized at the 1% and 99% level.

Table 4.5 shows that error companies pay higher audit fees (p-value < 0.05) than control companies in the subsequent two years, while there is no evidence that they are charged higher NAS fees over the same period. Regarding total fees, there is only weak evidence that error companies pay more than control companies (p-value < 0.10). Error companies purchase fewer NAS in the current period compared to control companies (p-value < 0.01). These results seem to be consistent with

⁵¹ A similar consideration is made in section 3.5.1 as well as in Fiallo/Hecker (2019a), p. 233 footnote no. 76.

⁵² As reported in section 3.5.1, Cram/Karan/Stuart (2009), pp. 494-495, suggest running the regression including an overall intercept and omitting one dummy pair. In this case, Cram/Karan/Stuart (2009), p. 495, caution that the results of an OLS on choice-based data cannot be generalised as “there exists no larger population to which the sample is similar”. The results should be corrected by reweighting the OLS by the inverse of the sampling rate in each stratum. The strata in the sample are among others specific to industry, year and accounting standard specific, which would make the collection of sampling information costly to implement. By not reweight the OLS, I caution that the results are valid only for the analysed sample. Conversely, when running a conditional logistic regression, this problem does not occur. See Cram/Karan/Stuart (2009), pp. 492-495.

the existence of economic bonding due to future quasi-rents and of knowledge spillovers generated by the provision of NAS in the current period. Error companies are less liquid (p-value < 0.01) and there is weak evidence that they are more leveraged (p-value < 0.10). In terms of performance, error companies are more likely to report losses (p-value < 0.01), but there is no indication that they have a significantly different *RoA* compared to control companies. Error companies are less likely to be audited by Big 4 auditors (p-value < 0.01) and more likely to have appointed a new auditor in the error year (p-value < 0.05)⁵³.

Table 4.5: Descriptive statistics

Variables	Error group						Control group						OLS/Clogit regression		
	Mean	Median	St. dev.	Min.	Max.	Obs. (m)	Mean	Median	St. dev.	Min.	Max.	Obs. (m)	Exp. sign	Coeff.	p-value
<i>Future_AF</i>	13.142	13.036	1.293	10.739	16.791	131	12.985	12.777	1.204	10.739	16.791	131	+	0.121**	0.018
<i>Future_NAF</i>	11.295	12.067	3.573	0.000	16.524	131	11.082	11.728	3.582	0.000	16.667	131	+	0.164	0.317
<i>Future_TF</i>	13.502	13.391	1.395	11.127	17.360	131	13.347	13.106	1.275	11.062	17.424	131	+	0.113*	0.053
<i>NASfees</i>	9.085	11.019	5.113	0.000	16.013	131	10.192	11.082	3.756	0.000	15.783	131	-	-1.186***	0.005
<i>RoA</i>	0.016	0.008	0.438	-0.481	4.756	131	0.018	0.026	0.120	-0.665	0.296	131	-	-0.001	0.494
<i>Loss</i>	0.397	0.000	0.491	0.000	1.000	131	0.229	0.000	0.422	0.000	1.000	131	+	0.835***	0.002
<i>Lev</i>	0.675	0.650	0.350	0.097	3.237	131	0.603	0.596	0.456	0.057	4.795	131	+	0.083*	0.052
<i>CFO</i>	0.017	0.018	0.112	-0.442	0.477	131	0.046	0.047	0.108	-0.427	0.377	131	-	-0.034***	0.005
<i>Growth</i>	0.292	0.057	1.239	-0.829	9.242	131	0.201	0.047	1.461	-1.000	16.092	131	+	0.084	0.308
<i>Big4</i>	0.504	1.000	0.502	0.000	1.000	131	0.672	1.000	0.471	0.000	1.000	131	-	-0.830***	0.003
<i>Change</i>	0.130	0.000	0.337	0.000	1.000	131	0.061	0.000	0.240	0.000	1.000	131	+	0.761**	0.039
<i>IFRSfirst</i>	0.092	0.000	0.290	0.000	1.000	131	0.084	0.000	0.278	0.000	1.000	131	+	0.062	0.451
<i>Size</i>	19.418	18.834	2.507	14.945	26.198	131	19.353	18.936	2.365	14.945	26.198	131	-/+	0.064	0.200

Notes: The model for the OLS is the following: $Var_{i,t} = \beta_0 + \beta_1 Error_{i,t} + \beta_2 Size_{i,t} + \sum_{y=1}^{m-1} \beta_y Pair_y$. For the dichotomous variables *Loss*, *Big4*, *Change* and *IFRSfirst* I run a conditional logistic regression with pairs identified as strata. The reported coefficients are for the variable *Error*. The reported p-values are two-sided. ***, **, * means significance at the 1%, 5% and 10% level.

⁵³ These results are similar to those obtained in section 3.5.1, except for the variables *RoA* and *Change*.

Table 4.6: Pearson's and Spearman's correlation coefficients (part 1)

Variables	Error	Future_AF	Future_NAF	Future_TF	NASfees	RoA	Loss
Error		0.0727	0.0781	0.0776	-0.0106	-0.2589*	0.2330*
Future_AF	0.0630		0.7784*	0.9686*	0.7467*	0.0696	-0.1331*
Future_NAF	0.0300	0.5478*		0.8930*	0.8206*	0.0589	-0.1272
Future_TF	0.0582	0.9766*	0.6406*		0.8037*	0.0805	-0.1496*
NASfees	-0.1229*	0.5472*	0.6553*	0.6057*		0.045	-0.1271
RoA	-0.0026	-0.0200	0.0052	-0.0203	0.0728		-0.7964*
Loss	0.1811*	-0.1311*	-0.0883	-0.1401*	-0.1319*	-0.3025*	
Lev	0.0885	0.1443*	-0.0193	0.1252*	0.0088	0.2044*	0.1826*
CFO	-0.1326*	0.1201	-0.0536	0.1112	0.0194	0.0872	-0.3744*
Growth	0.0339	-0.0932	0.0052	-0.0813	-0.0851	0.1120	-0.0389
Big4	-0.1706*	0.4363*	0.1936*	0.4362*	0.2817*	-0.0609	0.0134
Change	0.1169	-0.0244	0.0167	-0.0266	-0.1083	-0.0494	0.1170
IFRSfirst	0.0135	-0.0941	-0.0660	-0.1062	-0.1115	-0.0155	-0.0639
Size	0.0133	0.9101*	0.5439*	0.8996*	0.5395*	-0.0222	-0.1762*

Notes: The table reports the coefficients of Pearson's product-moment correlation (under the diagonal) and Spearman's rank-order correlation (above the diagonal). * means significance at 5% level or higher.

Tables 4.6 and 4.7 show the results of Pearson's and Spearman's correlation for the model variables⁵⁴. According to Pearson's r , the variable *Error* is negatively correlated with *NASfees*, which is consistent with the fact that error companies buy fewer NAS in the current period. *Error* is also negatively correlated with *CFO* and *Big4*⁵⁵, while it is positively correlated with *Loss*⁵⁶. As expected, strong correlations (> 0.05) exist among future fees and current NAS fees. According to Pearson's correlation, all fees measures (*Future_AF*, *Future_NAF*, *Future_TF*, *NASfees*) are strongly positively correlated with *Size*, indicating that bigger companies are likely to pay higher fees for both audit and non-audit services. Due to the existence of such strong correlations, I further calculate the Variance Inflation Factor (VIF)⁵⁷. The results (untabulated) show that all the VIFs are below the critical value of 10, indicating that the model might not suffer from multicollinearity⁵⁸.

⁵⁴ As discussed in section 3.5.1, Pearson's correlation coefficients can be interpreted even when one of the two variables is a category coded 0/1. This special case is known as point-biserial correlation. See e.g. Sheskin (2000), pp. 799-804, and Corder/Foreman (2009), pp. 134-142. In general, Pearson's correlation requires a linear relationship between the variables and for them to be numerical. Therefore, in Tables 4.6 and 4.7 are also reported Spearman's rank correlation coefficients, which can be calculated for ordinal variables. See e.g. Cleff (2019), pp. 94-95.

⁵⁵ *Big4* is significant at the 5% level according to Spearman's rank.

⁵⁶ *Loss* is significant at the 5% level according to Spearman's rank.

⁵⁷ Wooldridge (2016), p. 86.

⁵⁸ Other studies, which have used 10 as cut-off value, are e.g. Lopatta/Kaspereit/Canitz/Maas (2015), p. 580; Gros (2016), p. 235.

Table 4.7: Pearson's and Spearman's correlation coefficients (part 2)

Variables	Lev	CFO	Growth	Big4	Change	IFRSfirst	Size
Error	0.1741*	-0.1658*	0.036	-0.1483*	0.09	-0.0369	-0.0025
Future_AF	0.4157*	0.0246	-0.0123	0.4021*	0.0002	-0.0245	0.8931*
Future_NAF	0.3166*	-0.0383	-0.0799	0.3264*	0.0002	-0.0953	0.7635*
Future_TF	0.3996*	0.0233	-0.0315	0.4003*	-0.0029	-0.0486	0.8921*
NASfees	0.3014*	-0.0402	-0.0866	0.3384*	-0.1292	-0.0584	0.7360*
RoA	-0.2934*	0.4716*	0.3850*	-0.0364	-0.0907	0.0796	0.0753
Loss	0.1485*	-0.3375*	-0.2548*	-0.0022	0.1084	-0.0501	-0.1991*
Lev		-0.2223*	-0.09	0.1053	-0.0715	0.0482	0.3952*
CFO			0.1713*	0.0491	-0.0443	-0.1004	-0.0358
Growth				-0.0516	-0.0385	0.0667	-0.0064
Big4					-0.0295	-0.0335	0.4089*
Change						-0.0852	-0.038
IFRSfirst							-0.0222
Size							

Notes: The table reports the coefficients of Pearson's product-moment correlation (under the diagonal) and Spearman's rank-order correlation (above the diagonal). * means significance at 5% level or higher.

4.4.2 Multivariate results

Table 4.8 presents the results of the conditional logistic regression. According to the X^2 test, the models are all statistically significant (p-values < 0.01) and with an explanatory power ranging from 20% to 24%, which indicates a good fit, and which is line with previous studies⁵⁹. The coefficient of *Future_TF* (M3) is positive and statistically significant, indicating that fees paid in subsequent periods are positively associated with enforcement announcements. Similarly, *Blay/Geiger* (2013) find that future fees, also measured as the natural logarithm of the fees paid in the two subsequent periods, are negatively associated with the issue of a going concern opinion⁶⁰. Taken together, these findings are consistent with the existence of economic bonding, suggesting that the expected benefits from the future provision of audit services and NAS might impair auditor independence. *Blay/Geiger* (2013) do not consider future audit and non-audit fees separately, possibly assuming that it is the sum of all fees which determines the economic bond⁶¹. In this study, future audit and non-audit fees are also analysed separately as quasi-rents might be higher for non-audit services than for audit services⁶². The coefficient of *Future_AF* (M1) is positive and statistically significant, while there is only weak evidence of the variable *Future_NAF* being positively associated with the variable *Error* (M2). Also, the odds ratios suggest that the association

⁵⁹ E.g. Causholli/Chambers/Payne (2014), p. 691.

⁶⁰ Blay/Geiger (2013), p. 592.

⁶¹ Blay/Geiger (2013), p. 583.

⁶² E.g. Larcker/Richardson (2004), p. 632.

between future fees and audit quality might be driven by the level of audit fees. It is possible that the economic bond originates mainly from the provision of audit services rather than the provision of NAS. I offer the following interpretation. Since auditor independence is threatened by the expected quasi-rents, it is possible that audit services play a prominent role as the related quasi-rents are easier to predict and to evaluate, while the quasi-rents generated by NAS, especially if non-recurring, are more volatile. These results are only partially in line with those of *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018), who find that audit quality is likely to be affected by the expectations of future NAS purchases. Specifically, it is future fees for other non-audit services which are negatively associated with restatements and the propensity to meet or beat analysts' earnings forecasts, while there is no significant association with future tax and audit-related fees⁶³. Conversely, the study of *Fuhrmann* (2016) provides no evidence of an association between future NAS fees and audit quality⁶⁴.

Throughout M1 to M3, the coefficient of *NASfees* is negative and significant, indicating that the provision of NAS in the current period is associated with higher audit quality. This finding is consistent with the existence of knowledge spillovers from the provision of NAS, which might increase the ability of the auditor to detect error. Conversely, *Blay/Geiger* (2013) find that current NAS fees are negatively associated with the issue of a going concern opinion, which they interpret as a sign of economic bonding⁶⁵, while *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018) find no significant relation between restatements and current NAS fees and only weak evidence that current NAS fees are positively associated with the propensity to meet or beat analysts' earnings forecasts⁶⁶. Regarding the control variables, the existence of enforcement findings is positively associated with *Loss*, while there is weak evidence of a negative association between the variable *CFO* and enforcement findings⁶⁷. The results also show that error companies are less likely to be audited by a Big4 auditing company⁶⁸. Finally, there is some evidence that *Size* is associated with the existence of enforcement findings in M1 and M2⁶⁹.

⁶³ *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018), pp. 22-24.

⁶⁴ *Fuhrmann* (2016), p. 57.

⁶⁵ *Blay/Geiger* (2013), p. 598.

⁶⁶ *Castillo-Merino/Garcia-Blandon/Martinez-Blasco* (2018), pp. 22-24. Conversely, their results show that audit quality is negatively associated with current audit fees.

⁶⁷ The coefficient of *CFO* is significant at the 5% level in M1, not significant in M2, and significant at the 10% level in M3.

⁶⁸ For an interpretation, see section 3.5.2.1.

⁶⁹ In M2, the coefficient is only significant at the 10% level.

Table 4.8: Multivariate results

	expected sign	(M1)	(M1)	(M2)	(M2)	(M3)	(M3)
		coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio
<i>Future_AF</i>	+	1.010*** (0.003)	2.746***				
<i>Future_NAF</i>	+			0.097* (0.070)	1.102*		
<i>Future_TF</i>	+					0.901*** (0.002)	2.463***
<i>NASfees</i>	-	-0.130*** (0.004)	0.878***	-0.140*** (0.004)	0.869***	-0.155*** (0.001)	0.856***
<i>RoA</i>	-	1.025 (0.779)	2.786	1.303 (0.762)	3.681	1.172 (0.777)	3.230
<i>Loss</i>	+	1.045*** (0.008)	2.843***	1.073*** (0.009)	2.923***	1.136*** (0.006)	3.113***
<i>Lev</i>	+	-0.125 (0.614)	0.883	0.366 (0.211)	1.442	-0.033 (0.529)	0.968
<i>CFO</i>	-	-2.885** (0.043)	0.056**	-2.160 (0.103)	0.115	-2.412* (0.076)	0.090*
<i>Growth</i>	+	0.027 (0.401)	1.027	0.000 (0.499)	1.000	-0.002 (0.507)	0.998
<i>Big4</i>	-	-1.091*** (0.001)	0.336***	-0.897*** (0.005)	0.408***	-1.145*** (0.001)	0.318***
<i>Change</i>	+	0.544 (0.149)	1.724	0.563 (0.138)	1.756	0.540 (0.155)	1.716
<i>IFRSfirst</i>	+	-0.081 (0.558)	0.922	0.003 (0.498)	1.003	0.108 (0.423)	1.114
<i>Size</i>	-/+	0.158** (0.016)	1.172**	0.681* (0.093)	1.975*	0.181 (0.686)	1.198
Obs.		262		262		262	
Prob > X ²		0.000		0.000		0.000	
Pseudo R ²		0.239		0.203		0.244	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff.}$ and indicate the percentage change in the probability of the outcome occurring (the company having an enforcement announcement) versus the outcome not occurring, by a 1-unit increase in the independent variable. In the case of dichotomous predictors, they indicate the increase or decrease in the odds by the occurrence of the effect coded in the categorical variable. For more considerations on odds ratios, see Hosmer/Lemeshow/Sturdivant (2013), p. 50-52, 63. Prob > X² indicates at which level it is possible to reject the null hypothesis that the predictors have no effect on the dependent variable. The goodness of fit is expressed by the McFadden's pseudo R².

4.4.3 Sensitivity tests

4.4.3.1 Financial companies

In order to assess the robustness of the findings, I perform additional tests. As mentioned previously, one fourth of the sample (66 obs.) is made up of banks, insurance and real estate companies. These companies are now dropped from the sample, while the regressions on the models M1, M2 and M3 are rerun⁷⁰. Table 4.9 shows that the models are still significant and possess a higher explanatory power (32%-38.5%) compared to the models run on the full sample (20%-24%). As in the main analyses, audit quality is negatively associated with the level of future audit fees, future NAS fees and future total fees⁷¹ and positively associated with current NAS. I conclude that the evidence obtained in the main model is robust when excluding financial companies.

4.4.3.2 Auditor change

So far, the expected value of quasi-rents is measured with future fees, regardless if they are charged by the incumbent auditor or by a new auditor. The assumption here is that in the current period the incumbent auditor might not know that he will not be reappointed over the following two years. However, the fees charged by the new auditor might differ from those which the incumbent auditor would have expected. A discrepancy between actual fees and expected fees might have a confounding effect on the fee variables and bias the results.

The regressions for the models M1 to M3 are rerun, dropping from the sample all firm/year observations in which the incumbent auditor was not reappointed for the subsequent two periods. There remain a sample of 184 observations. Table 4.10 shows that all models are significant and have sufficient explanatory power (26%-29%). The coefficients of the fee variables are consistent in terms of sign and significance with those obtained so far, which indicate that the subsequent auditor change does not alter the results.

⁷⁰ For the reasoning behind this sensitivity test, see section 3.5.3.1.

⁷¹ Again, the coefficient of *Future_NAF* is only significant at the 10% level.

Table 4.9: Multivariate results - without financial companies

	expected sign	(M1)	(M1)	(M2)	(M2)	(M3)	(M3)
		coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio
<i>Future_AF</i>	+	1.022** (0.011)	2.779**				
<i>Future_NAF</i>	+			0.164* (0.052)	1.178*		
<i>Future_TF</i>	+					1.368*** (0.001)	3.927***
<i>NASfees</i>	-	-0.222*** (0.002)	0.801***	-0.263*** (0.003)	0.769***	-0.306*** (0.001)	0.736***
<i>RoA</i>	-	0.442 (0.671)	1.555	-0.107 (0.466)	0.899	0.571 (0.708)	1.770
<i>Loss</i>	+	1.034** (0.033)	2.811**	0.944** (0.047)	2.570**	1.223** (0.022)	3.399**
<i>Lev</i>	+	-0.003 (0.503)	0.997	0.547 (0.114)	1.728	0.040 (0.465)	1.041
<i>CFO</i>	-	-6.211** (0.010)	0.002**	-5.862** (0.017)	0.003**	-6.127** (0.014)	0.002**
<i>Growth</i>	+	0.006 (0.483)	1.006	-0.001 (0.502)	0.999	-0.018 (0.549)	0.982
<i>Big4</i>	-	-1.495*** (0.002)	0.224***	-1.377*** (0.003)	0.252***	-1.754*** (0.001)	0.173***
<i>Change</i>	+	0.542 (0.253)	1.719	0.396 (0.316)	1.486	0.652 (0.226)	1.919
<i>IFRSfirst</i>	+	-0.268 (0.640)	0.765	-0.103 (0.552)	0.902	0.036 (0.481)	1.037
<i>Size</i>	-/+	-0.047* (0.067)	0.954*	0.342 (0.427)	1.408	-0.328 (0.532)	0.720
Obs.		196		196		196	
Prob > X ²		0.000		0.000		0.000	
Pseudo R ²		0.345		0.323		0.385	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff.}$. The analysis is based on a restricted sample of 196 observations.

Table 4.10: Multivariate results - without subsequent auditor changes and with client importance

	expected sign	(M1)		(M2)		(M3)		(M4)	
		coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio
<i>Future_AF</i>	+	0.856** (0.025)	2.353**						
<i>Future_NAF</i>	+			0.125* (0.098)	1.133*				
<i>Future_TF</i>	+					0.893*** (0.010)	2.443***		
<i>Future_CI</i>	+							-1.218 (0.637)	0.296
<i>NASfees</i>	-	-0.135** (0.021)	0.874**	-0.142** (0.024)	0.868**	-0.159** (0.013)	0.853**	-0.106* (0.072)	0.899*
<i>RoA</i>	-	1.783 (0.765)	5.948	2.172 (0.801)	8.780	2.085 (0.793)	8.047	0.961 (0.646)	2.615
<i>Loss</i>	+	1.346** (0.017)	3.842**	1.390** (0.014)	4.014**	1.491** (0.012)	4.443**	1.604*** (0.008)	4.975***
<i>Lev</i>	+	-0.009 (0.507)	0.991	0.499 (0.173)	1.647	0.096 (0.425)	1.100	0.267 (0.283)	1.307
<i>CFO</i>	-	-4.030** (0.044)	0.018**	-3.326* (0.082)	0.036*	-3.799* (0.053)	0.022*	-4.048* (0.072)	0.017*
<i>Growth</i>	+	-0.009 (0.527)	0.991	-0.028 (0.588)	0.972	-0.027 (0.583)	0.973	-0.046 (0.624)	0.955
<i>Big4</i>	-	-0.869** (0.023)	0.420**	-0.696** (0.049)	0.499**	-1.012** (0.013)	0.363**	-0.958** (0.028)	0.384**
<i>Change</i>	+	0.589 (0.155)	1.802	0.650 (0.128)	1.915	0.578 (0.168)	1.782	0.508 (0.207)	1.661
<i>IFRSfirst</i>	+	0.174 (0.405)	1.190	0.270 (0.357)	1.310	0.301 (0.338)	1.351	0.228 (0.405)	1.256
<i>Size</i>	-/+	1.030** (0.033)	2.800**	1.564** (0.013)	4.777**	1.111* (0.088)	3.036*	1.771** (0.043)	5.878**
Obs.		184		184		184		162	
Prob > X ²		0.000		0.001		0.000		0.000	
Pseudo R ²		0.275		0.256		0.288		0.299	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff.}$. The analysis is based on a restricted sample of 184 and 162 observations.

4.4.3.3 Client importance

So far, this study focused on whether the value of quasi-rents an auditor expects from a client affects audit quality, without considering that the auditor also expects quasi-rents from other clients. *DeAngelo* (1981a, 1981b) argue that auditor decisions result from the cost-benefit consideration of keeping the quasi-rents expected from a client against the probability of being discovered “cheating” and losing the quasi-rents expected from other clients⁷². It follows that when a client is economically important, the auditor might impair his independence to keep the client. Similarly, the European Regulator has recognised that excessive dependence on a client might be a threat to auditor independence⁷³, which is why I expect a negative association between audit quality and expected client importance

The economic importance of a client is represented by the ratio of the client-specific quasi-rents to the quasi-rents from all clients⁷⁴. Assuming again a positive correlation between quasi-rents and fees, I approximate client importance with the ratio of total fees to the auditor’s revenue⁷⁵. The variable *Future_CI* is measured as the average total fees paid to the incumbent auditor in the two subsequent years scaled by the average auditor’s revenue also in the two subsequent years. Revenue data are collected from auditor’s transparency reports, where auditors are required to disclose their revenues⁷⁶. The full sample (262 obs.) must be reduced due to missing revenue data⁷⁷. Additionally, all observations in which the incumbent auditor was not reappointed for the subsequent two periods are eliminated. There remains a sample of 162 observations.

Table 4.10 (M4) shows that the coefficient of *Future_CI* is not significant, indicating the lack of association between future client importance and audit quality. I offer here two possible interpretations: one, as the companies in the sample on average contribute to less than 3% of an auditor’s revenue in the two subsequent years, it is possible that by such a “low” client importance,

⁷² *DeAngelo* (1981a), p. 117; *DeAngelo* (1981b), p.184. Similar considerations are included in *Fiallo/Hecker* (2019b), p. 296.

⁷³ Recital 7 Regulation (EU) No 537/2014.

⁷⁴ *DeAngelo* (1981b), pp. 187-192.

⁷⁵ E.g. *Ostrowski/Söder* (1999), p. 558; *Chung/Kallapur* (2003), p. 932.

⁷⁶ For a more detailed description on collecting revenue data as well as on the assumptions made to identify the relevant transparency reports, see *Fiallo/Hecker* (2019a), pp. 228-229 and *Fiallo/Hecker* (2019b), pp. 294-295.

⁷⁷ E.g. transparency reports are available at the earliest for client’s financial statements beginning in 2006. For further details on data collection as well as references to the regulatory framework, see *Fiallo/Hecker* (2019b), pp. 294-295.

auditors might not be willing to compromise their independence⁷⁸; two, measuring client importance at a firm level might not capture the economic dependence between client and auditor. Prior studies suggest that short-term profitability might be more relevant to partners than a firm's reputation⁷⁹, which is why when data are available client importance might be measured at office or at partner level.

4.4.3.4 Time frame

So far, the analyses focused on the fees paid in the two subsequent years. As discussed above, a two-year window might be appropriate due to the comparatively higher evaluation of expected benefits in the near future and to the fact that the auditor's expectations tend to diverge from the actual fees over time⁸⁰. As a last sensitivity test, I verify whether the findings are sensitive to the chosen time-span, by employing a shorter period (one year) and a longer period (three years)⁸¹. Multiple regressions are run on models M1 to M3, using alternatively the natural log of the fees paid in the next year and the natural log of the fees paid in the three subsequent years.

The results are reported in Table 4.11 and 4.12. While the results obtained so far provide only weak evidence that future fees for NAS are associated with the existence of enforcement findings, the sensitivity test shows that audit quality is negatively associated with the fees paid in the following year for NAS⁸². This might be consistent with the quasi-rents for NAS to be more volatile, with the auditor being able to predict them correctly only in the following period. The other coefficients of the fee variables are, in terms of sign and significance, qualitatively similar to the results obtained so far, which means the findings are robust to the chosen time-span.

⁷⁸ The percentage is considered low if compared to the 15% threshold set by art. 4 para. 3 Regulation (EU) No 537/2014 is 15%. For a detailed analysis of the EU-regulation, including the comparison with the (prior) national regulation, which also imposed a 15% limit, see Fiallo/Hecker (2019b), p. 294.

⁷⁹ E.g. Causholli/Chambers/Payne (2014), pp. 682-683.

⁸⁰ See section 4.3.2.

⁸¹ This sensitivity test is suggested by e.g. Blay/Geiger (2013), p. 598.

⁸² The coefficient of *Future_NAF* is significant at the 5% level, when using the log of the fees paid in the following year, while it is only significant at the 10% level when considering a 2-year or a 3-year period.

Table 4.11: Multivariate results - one-year period

	expected sign	(M1)	(M1)	(M2)	(M2)	(M3)	(M3)
		coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio
<i>Future_AF (1y)</i>	+	0.964*** (0.003)	2.623***				
<i>Future_NAF (1y)</i>	+			0.099** (0.036)	1.104**		
<i>Future_TF (1y)</i>	+					0.900*** (0.001)	2.458***
<i>NASfees</i>	-	-0.135*** (0.003)	0.874***	-0.157*** (0.003)	0.855***	-0.170*** (0.001)	0.844***
<i>RoA</i>	-	1.025 (0.766)	2.787	1.298 (0.764)	3.662	1.223 (0.784)	3.399
<i>Loss</i>	+	1.027*** (0.010)	2.793***	1.127*** (0.007)	3.088***	1.152*** (0.006)	3.164***
<i>Lev</i>	+	-0.170 (0.647)	0.843	0.226 (0.299)	1.254	-0.175 (0.646)	0.839
<i>CFO</i>	-	-3.453** (0.023)	0.032**	-2.176 (0.102)	0.113	-2.861** (0.046)	0.057**
<i>Growth</i>	+	0.016 (0.442)	1.016	0.040 (0.358)	1.041	0.000 (0.499)	1.000
<i>Big4</i>	-	-1.050*** (0.002)	0.350***	-0.928*** (0.004)	0.395***	-1.074*** (0.002)	0.342***
<i>Change</i>	+	0.554 (0.142)	1.741	0.628 (0.112)	1.874	0.531 (0.159)	1.700
<i>IFRSfirst</i>	+	-0.132 (0.592)	0.877	-0.043 (0.531)	0.957	0.160 (0.390)	1.173
<i>Size</i>	+/-	0.249** (0.019)	1.282**	0.678* (0.093)	1.970*	0.205 (0.642)	1.227
Obs.		262		262		262	
Prob > X ²		0.000		0.000		0.000	
Pseudo R ²		0.238		0.210		0.248	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff.}$. The variables *Future_AF (1y)*, *Future_NAF (1y)* and *Future_TF (1y)* are computed as the natural log of the fees paid in the next year.

Table 4.12: Multivariate results - three-year period

	expected sign	(M1)	(M1)	(M2)	(M2)	(M3)	(M3)
		coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio	coeff. (p-value)	odds ratio
<i>Future_AF (3y)</i>	+	0.888*** (0.007)	2.429***				
<i>Future_NAF (3y)</i>	+			0.095* (0.096)	1.099*		
<i>Future_TF (3y)</i>	+					0.762*** (0.008)	2.142***
<i>NASfees</i>	-	-0.108** (0.013)	0.897**	-0.123** (0.011)	0.884**	-0.127*** (0.007)	0.881***
<i>RoA</i>	-	0.655 (0.741)	1.925	0.597 (0.655)	1.818	0.711 (0.738)	2.037
<i>Loss</i>	+	1.177*** (0.003)	3.246***	1.186*** (0.004)	3.275***	1.222*** (0.003)	3.396***
<i>Lev</i>	+	-0.013 (0.513)	0.987	0.201 (0.313)	1.222	0.037 (0.465)	1.038
<i>CFO</i>	-	-2.079 (0.110)	0.125	-1.800 (0.153)	0.165	-1.918 (0.131)	0.147
<i>Growth</i>	+	0.015 (0.450)	1.015	-0.014 (0.547)	0.986	-0.004 (0.513)	0.996
<i>Big4</i>	-	-1.127*** (0.002)	0.324***	-0.968*** (0.004)	0.380***	-1.127*** (0.002)	0.324***
<i>Change</i>	+	0.581 (0.139)	1.787	0.641 (0.111)	1.899	0.555 (0.154)	1.742
<i>IFRSfirst</i>	+	-0.304 (0.697)	0.738	-0.261 (0.668)	0.771	-0.190 (0.625)	0.827
<i>Size</i>	-/+	0.113*** (0.007)	1.119***	0.557 (0.157)	1.746	0.130 (0.769)	1.139
Obs.		236		236		236	
Prob > X ²		0.000		0.000		0.000	
Pseudo R ²		0.241		0.210		0.240	

Notes: ***, **, * means significance at the 1%, 5% and 10% level. When I formulate predictions regarding the sign of the coefficient, the reported p-values are from a one-sided test, otherwise from a two-sided test. The reported odds ratios are computed as $OR = e^{coeff.}$. The variables *Future_AF (3y)*, *Future_NAF (3y)* and *Future_TF (3y)* are computed as the natural log of the fees paid in the three subsequent years. The analysis is based on a restricted sample of 236.

4.5 Conclusion

In this paper, I examine how the current provision of NAS as well as the expectations regarding the future provision of audit and non-audit services affect audit quality. This multi-period approach is necessary as auditor independence might be negatively affected by the expectations of future quasi-rents⁸³, while the existence of knowledge spillovers in the current period might improve the auditor's ability to detect errors⁸⁴.

In line with the study of *Blay/Geiger (2013)*⁸⁵, I observe that the presence of enforcement findings is positively associated with future fees, which is consistent with the existence of economic bonding. This seems to suggest that the expected benefits from the future provision of audit services and NAS might impair auditor independence and lead to lower audit quality. Interestingly, the analysis shows that the association between future fees and audit quality seems to be driven by the level of future audit fees, while the results provide limited support for future NAS fees being negatively associated with audit quality⁸⁶. These findings suggest that the economic bond might originate from the provision of audit services rather than the provision of NAS, as the audit-related quasi-rents are might be easier to predict and to evaluate, while the NAS-related quasi-rents might be more volatile. Regarding the current provision of NAS, contrary to what was reported by *Blay/Geiger (2013)*⁸⁷ and *Castillo-Merino/Garcia-Blandon/Martinez-Blasco (2018)*⁸⁸, I find evidence consistent with the presence of knowledge spillovers, which improve the ability of the auditor to detect errors. This is in line with the findings discussed in section 3.5.2.1 as well as in *Fiallo/Hecker (2019a)*⁸⁹, where there are indications that certain NAS, e.g. tax services, might improve audit quality.

The sensitivity tests show that the results are robust to different time frames (one, two and three years) and are not affected by the presence of companies from the financial sector. To make sure

⁸³ DeAngelo (1981a), pp. 118-123.

⁸⁴ On the potential benefits of knowledge spillovers on audit quality see Fiallo/Hecker (2019a), p. 231. A more detailed analysis of the theoretical arguments beyond knowledge spillovers is included in section 3.3.1.2 of this study.

⁸⁵ Blay/Geiger (2013), p. 592.

⁸⁶ These results are only partially in line with those of Castillo-Merino/Garcia-Blandon/Martinez-Blasco (2018), p. 22-24, who find that audit quality is likely to be affected by the expectations of future purchase of other non-audit services, but not by the future provision of tax and audit-related services.

⁸⁷ Blay/Geiger (2013), p. 598.

⁸⁸ Castillo-Merino/Garcia-Blandon/Martinez-Blasco (2018), pp. 22-24.

⁸⁹ Fiallo/Hecker (2019a), pp. 232-234.

the results are not distorted by subsequent auditor changes, all firm/year observations, where the incumbent auditor was not reappointed in the two subsequent years, are excluded from the sample, as the actual fees paid to another auditor might differ from those which would have been charged or expected by the incumbent auditor. The results of this test are consistent with the findings obtained in the main analysis. Another aspect considered, is that the risk of economic bonding might also depend on client importance⁹⁰. However, the results do not show that the expected client importance is significantly associated with the existence of enforcement announcements.

In conclusion, this study helps to provide a better understanding of the relation between audit quality and the provision of NAS in a multi-period framework. It also provides some answers to the concerns expressed recently by the European regulator regarding the provision of certain NAS⁹¹ and the level and structure of the fees paid by audited companies to their statutory auditor⁹². The results of this study indicate that the provision of NAS might generate knowledge spillovers, which increase the ability of the auditor to detect errors. Considering this evidence, there seems to be no need for further limiting the provision of NAS, as this might reduce audit quality. Also, there is only limited evidence that the expected provision of future NAS is associated with enforcement findings, so that it is unclear whether reducing the possibility to provide NAS or putting a cap on NAS fees might help to increase auditor independence.

This study suffers from some limitations⁹³. As audit quality cannot be directly measured, I employ as a surrogate the existence of enforcement findings. The conclusions drawn from this analysis depend on how well this proxy reflects actual audit quality. Also, they depend on the research design and the choice of the control companies and on the reliability of the underlying data. Prior studies express caution about the existence of a fee bias e.g. due to the inaccurate disclosure of fees by the audited entities⁹⁴ and due to ambiguities in the disclosure requirements, which might have led to inconsistencies in practice⁹⁵. As this study is set in Germany before the introduction of the

⁹⁰ The concern of an auditor becoming excessively dependent on a client has been expressed by the EU regulator, see Recital 7 Regulation (EU) No 537/2014.

⁹¹ Recital 8 Regulation (EU) No 537/2014. For a detailed analysis of certain aspects of Regulation (EU) No 537/2014, including the comparison with the (prior) national regulation in Germany, see Fiallo/Hecker (2019a), p. 230 and Fiallo/Hecker (2019b), pp. 290-291, 294.

⁹² Recital 7 Regulation (EU) No 537/2014.

⁹³ For a discussion of the limitations, see Fiallo/Hecker (2019c), p. 348 as well as section 3.6.

⁹⁴ E.g. Markelevich/Rosner (2013), p. 1621; Fiallo/Hecker (2019a), p. 231.

⁹⁵ E.g. Köhler/Ratzinger-Sakel (2012), p. 286; Quick/Sattler (2011a), p. 78; Sattler (2011), pp. 325-326; Krauß/Zülch (2013), p. 322.

Regulation (EU) No 537/2014, the results might not apply to other countries and should be interpreted considering the existing national regulation between 2005 and 2017.

Finally, a typical problem of inferences in audit research is that audit fees, NAS fees and accounting quality might be endogenous variables, due to common factors affecting simultaneously the choice of services acquired by a company and the quality of a company's financial reporting⁹⁶. The results obtained indicate an association between future fees and enforcement findings. However, the causality of this relation remains unclear⁹⁷. E.g. *Laschewski/Möller/Risse* (2014) find that audit fees, fees for other assurance services and other non-audit services are likely to increase in the two years following the erroneous financial statement. However, they suggest that this increase is consistent with additional services provided by the auditor during and after the enforcement examination and with the pricing of the higher risk related to the discovery of erroneous accounting⁹⁸. Therefore, it cannot be ruled out that the positive association between future fees and enforcement findings is to some extent due to the higher risk and effort. Notwithstanding these limitations, the results obtained in this study are consistent with the quasi-rent theory.

⁹⁶ Antle/Gordon/Narayanamoorthy/Zhou (2006), pp. 236-237; Markelevich/Rosner (2013), pp. 1612-1613.

⁹⁷ A similar concern is expressed by Blay/Geiger (2013), p. 602, when discussing the association between future fees and the issuing of a going concern opinion.

⁹⁸ *Laschewski/Möller/Risse* (2014), p. 312.

5. Earnings management and the German enforcement system: An application of the m-score

Abstract: The two-tier German enforcement system was introduced in 2004 to enhance the quality of financial statements and restore market confidence. The efficiency of the enforcement system depends also on its ability to identify companies manipulating their earnings, which is why it is important to provide evidence about the extent of earnings management among companies falling under the enforcement regime. Also, the nature of enforcement findings is fundamental for determining the future actions aimed at limiting erroneous accounting. In this study, I analyse the relation between erroneous financial statements, identified with enforcement announcements, and earnings management, using the probability of manipulation (m-score). In line with prior studies, the results suggest that the German enforcement system might be efficient in finding companies engaging in earnings management. By employing several financial statement ratios, I am able to shed some light on the determinants of erroneous financial reporting. The results show that financial pressure is associated with erroneous financial reporting and that earnings management might take place through the manipulation of sales. Finally, the analysis of the information content of enforcement announcements does not show that error severity and lack of cooperation of the company's managers are indications of ongoing earnings management.

5.1 Introduction

The two-tier German enforcement system was introduced in 2004 with the Financial Reporting Compliance Act (Bilanzkontrollgesetz) to enhance the quality of financial statements and restore market confidence after some accounting scandals¹. The Financial Reporting Enforcement Panel (FREP) and the German Federal Financial Supervisory Authority (BaFin) examine the financial reporting of publicly listed companies. If errors are identified during the examination, an enforcement announcement must be disclosed. The choice of a “name and shame” mechanism aims at sanctioning companies with erroneous accounting and at increasing the transparency of financial information². The enforcement institutions carried out more than 1300 examinations between 2005 and 2017, which ended with enforcement announcements in 258 cases (19.8%)³. The FREP provides only limited information about the determinants of accounting errors⁴. In fact, enforcement announcements do not clarify whether errors are unintentional or the result of manipulations, failing to inform about the extent of earnings management among the companies subject to enforcement findings nor about the ability of the enforcement institutions to detect manipulators⁵.

The nature of the enforcement findings is fundamental for determining the future actions aimed at limiting erroneous accounting⁶. While unintentional mistakes might be due to GAAP complexity, weak control systems, lack of resources or expertise⁷, earnings management involves the voluntary adjustment of financial information to provide a certain portrait of a company⁸. Also, the ability of the enforcement system to enhance the quality of financial statements, depends on its capacity

¹ CESR Standard no. 1 on financial information - enforcement of standards on financial information in Europe, pp. 3-4.

² For more details on the functioning of the German enforcement system, see section 5.2.1.

³ Thormann/Barth (2018), p. 474.

⁴ Over the years, some of the causes of errors identified by the FREP have been e.g.: complexity of IFRS, complicated transactions, lack of knowledge of accounting standards, the financial crisis, the economic situation of the company and the conduct of earnings management. See FREP (2018a), p. 6; FREP (2017), p. 7; FREP (2016), p. 6; FREP (2015a), p. 6; FREP (2014), p. 7; FREP (2013), p. 7; FREP (2012), p. 8; FREP (2011), p. 5; FREP (2010), appendices p. 7; FREP (2009), appendices p. 9; FREP (2008), appendices p. 6.

⁵ In this study, I use the wording “earnings management” and “earnings manipulation” as synonyms. In the literature, the term “fraud” is frequently used to indicate earnings manipulation. The Oxford Dictionary defines fraud as “wrongful or criminal deception intended to result in financial or personal gain”. From this definition follows that the concept of fraud is broader than that of earnings manipulation. In this paper, I will use the term fraud only when describing the work of other authors, otherwise I will use earnings management or manipulation.

⁶ Pasch (2017), p. 3.

⁷ E.g. Ernstberger/Hitz/Stich (2012), p. 4.

⁸ Healy/Wahlen (1999), p. 368; Wagenhofer/Ewert (2015), p. 265.

to identify companies engaging in earnings management⁹, which is why it is important to provide evidence about the extent of earnings management among companies falling under the enforcement regime.

In this paper, I analyse the relation between erroneous financial statements, identified by enforcement announcements, and the presence of earnings management. I test whether companies with enforcement findings have, in the error year, a higher probability of manipulation than companies subject to the enforcement regime, but which have not issued any enforcement announcements. The extent to which accounting errors, discovered during enforcement examinations, are due to intentional misstatements, is estimated using the probability of manipulation (m-score)¹⁰. Additionally, I test whether the existence of enforcement announcements can be predicted by using different financial statement ratios. Finally, I focus on the information content of announcements, in terms of error severity and managers' lack of cooperation, to examine whether they might indicate a higher probability of manipulation.

This paper makes several contributions. First, it contributes to the debate on the efficiency of the German enforcement system to enhance accounting quality and to expose companies which manipulate their financial statements. Prior studies, e.g. *Böcking/Gros/Worret (2015)*, *Strohmenger (2013)*, *Strohmenger (2014)* and *Hoehn/Strohmenger (2013)*, find evidence of earnings management among companies subject to enforcement announcements. I contribute to this field of research by analysing the relation between enforcement announcements and earnings management using the m-score. Prior studies have mostly employed discretionary accruals to proxy for the existence of earnings management. Even though discretionary accruals are broadly used in accounting research, they have received much criticism¹¹. Second, it advances the knowledge on the causes of erroneous accounting. Acknowledging the evidence of *Pasch (2017)* and *Strohmenger (2014)*, who suggest that certain financial statement ratios might be more useful in explaining the presence of errors in companies subject to the enforcement regime than discretionary

⁹ Böcking/Gros/Worret (2015), p. 433; Strohmenger (2014), pp. 273-274.

¹⁰ Building on the ratio analysis and the explanatory power of financial statement ratios, Beneish (1999), pp. 24-26, developed a series of indices, based on financial statement data, that added together give the probability of manipulation (m-score) for a certain company in a given year.

¹¹ E.g. Jackson (2018), pp. 137-138, criticises the use of discretionary accrual models which estimate the level of "normal" accruals using a cross-sectional approach where companies are clustered by sector and year. He argues that "econometrically speaking, discretionary accruals are simply deviations from industry averages", while attributing this difference entirely to earnings management is an overconfident assumption.

accruals¹², this paper uses several financial statement ratios to predict the existence of enforcement announcements. Third, by analysing the information reported on the enforcement announcements, it provides information on how to interpret the content of announcements in relation to earnings management.

The research sample includes 2561 firm/year observations regarding non-financial companies. Companies subject to enforcement announcements from July 2005 to December 2017 form the error group (121 firm/year observations), while companies falling under the enforcement regime, which do not have any enforcement announcements as of 31 December 2017 are the non-error group (2440 firm/year observations).

The univariate results show that companies with enforcement announcements have a higher probability of manipulation in the error year than companies subject to the enforcement regime but without enforcement announcements. These results are in line with prior studies, which determine that the German enforcement system is efficient in finding companies which engage in earnings management¹³. By employing several financial statements ratios, I find some indications that financial pressure is associated with erroneous financial reporting and that earnings management might take place through the manipulation of sales, while there is only weak evidence that it is conducted with the improper capitalisation of non-current assets or that an increase in the cost of sales might be the trigger. Finally, the analysis of the content of enforcement announcements is inconclusive, not providing any evidence that companies with more severe errors or uncooperative companies are more likely to manipulate their earnings.

The remainder of this paper is structured as follows. Section 5.2 describes the institutional framework focusing on the functioning of the enforcement system and the work of the FREP. In section 5.3, I outline the theory of earnings management before formulating the research hypotheses. Section 5.4 describes the m-score as developed by *Beneish* (1999), which is employed in this paper to measure the probability of manipulation. The research design is presented in section 5.5 and the results are discussed in section 5.6. Finally, the conclusions are presented in section 5.7.

¹² Pasch (2017), pp. 27-28; Strohmenger (2014), p. 273.

¹³ Böcking/Gros/Worret (2015), p. 482, Strohmenger (2013), p. 28, Strohmenger (2014), p. 296, and Hoehn/Strohmenger (2013), p. 3.

5.2 Institutional framework

5.2.1 The German enforcement system

The two-tier enforcement system was introduced in 2004 under the Financial Reporting Compliance Act (Bilanzkontrollgesetz)¹⁴ to prevent and sanction the erroneous financial statements of companies with securities admitted to trading on a regulated segment of the national market and with Germany as their home country¹⁵. Subject to the enforcement control are the most recent available (consolidated) annual financial statements with the respective (group) management report. If there are concrete indications of errors or upon request of the BaFin also interim financial reports with the related interim management report, (consolidated) payment reports and prior year financial statements and reports may be subject to an examination¹⁶.

The enforcement control is carried out on two levels. At the first level, the FREP conducts the examinations. The FREP is a private body and companies can choose whether to collaborate with the examination or not. At the second level, the BaFin holds the authority to enforce the participation of the company. Usually, the entire enforcement process is carried out by the FREP, which then notifies the BaFin of its findings. The BaFin acts when the company refuses (or withdraws) the collaboration with the FREP, the company does not accept the FREP's findings, or if concerns exist about how the examination was conducted by the FREP¹⁷.

According to sec. 342b para. 2 sentence 3 HGB, examinations are either reactive or proactive. Reactive examinations are initiated when there is concrete evidence of the misapplication of

¹⁴ The German enforcement system is currently regulated by sec. 342b-342e of the German Commercial Code (HGB) and sec. 106-113 of the German Securities Trading Act (Wertpapierhandelsgesetz - WpHG). For a description of the enforcement system, see e.g. Hecker/Wild (2012), pp. 5-8.

¹⁵ Sec. 342b para. 2 HGB. Since 1 January 2016, foreign companies with securities listed on the German regulated market are no longer subject to the enforcement regime. This change was introduced by the Act on the Implementation of the Directive Amending the Transparency Directive (Gesetz zur Umsetzung der Transparenzrichtlinie-Änderungsrichtlinie) to avoid having companies falling under multiple enforcement regimes in different countries. On the issue of multiple enforcements, see e.g. the discussion of Bockmann (2012), p. 158.

¹⁶ Sec. 342b paras. 2 and 2a HGB. The possibility to examine interim reports was introduced in 2007. Since 1 January 2016, the Act on the Implementation of the Directive Amending the Transparency Directive (Gesetz zur Umsetzung der Transparenzrichtlinie-Änderungsrichtlinie) allows prior year financial statements and (consolidated) transparency reports to be subject to examination. For a detailed analysis of the changes brought by the Act to the enforcement system, see e.g. Diemers/Homfeldt (2016), pp. 48-49.

¹⁷ Sec. 108 para. 1 WpHG. Sec 108 para. 1 sentence 4 WpHG allows the BaFin to assume the lead in the examination if the company is a credit institute, an insurance or an investment company, which is already under a supervisory investigation in the same matter.

accounting standards (indication-based examinations)¹⁸ or upon request by the BaFin (request-based examinations). Conversely, proactive examinations are random and follow a three-step sampling approach¹⁹, which combines: a risk-based selection, where companies are chosen if they display some risk factors (e.g. first-time listing, mergers and acquisitions, related party transactions, dismissal or resignation of one or more members of the board of directors or the supervisory board)²⁰; a stratified sampling selection, aiming at examining companies listed on the DAX, MDAX, SDAX and TecDAX every 4 to 5 years and all other companies every 8 to 10 years; an additional random selection, which makes sure that any company can be selected at any time²¹.

Examinations focus on so called enforcement priorities, which are disclosed each year in advance by the FREP²². Enforcement priorities are accounting issues that proved to be erroneous in the previous years or that are likely to generate errors in the future²³. An enforcement examination is therefore not a repetition of the statutory audit²⁴. In the case of reactive examinations, the focus is usually restricted to the matters that prompted the examination, unless further indications of erroneous accounting are found during the process²⁵.

If at the end of the examination, financial statements are determined to be erroneous and this information is relevant to the financial market, the company is required to publish an enforcement announcement in the Federal Gazette (Bundesanzeiger)²⁶. When an error is relevant to the market is not explicitly defined. According to the CESR standard no. 1 on financial information,

¹⁸ Indications might be provided e.g. by the press or by a third party, see FREP (2018a), p. 13.

¹⁹ The approach was revised in 2016 and again in 2018, see DPR (2016), pp. 1-5 and DPR (2018), pp. 1-4. Prior to that, companies were sampled according to a two-step sampling approach: partially risk-based (30%) and partially random (70%). For a description of the sampling technique before 2016, see DPR (2005), pp. 1-4 and DPR (2009), pp. 1-4.

²⁰ Thormann/Barth (2018), p. 482.

²¹ E.g. FREP (2017), pp. 19-20.

²² The lists with the enforcement priorities identified in each year are available at:

http://www.frep.info/pruefverfahren/pruefungsschwerpunkte_en.php

²³ The enforcement priorities for 2018 were: disclosure of the expected impact of implementation of major new standards in the period of their initial application; specific recognition, measurement, and disclosure issues of IFRS 3 Business Combinations; specific issues of IAS 7 Statement of Cash Flows; recognition and measurement of provisions in accordance with IAS 37 and related disclosures; group management report and group declarations, see FREP (2018a), p. 16.

²⁴ E.g. Ernst/Amann/Krais (2017), p. 238; Bockmann (2012), p. 163. For a detailed comparison between the statutory audit and the enforcement examination, see Beyhs/Kühne/Zülch (2012), pp. 650-660.

²⁵ FREP (2018b), p. 5.

²⁶ As well as in a multi-regional financial newspaper or an additional electronic financial information database. See sec. 109 para. 2 WpHG as well as e.g. Hecker/Wild (2012), pp. 6-7.

“materiality should be assessed according to the relevant reporting framework”²⁷. Under IFRS, information is material when it can influence the decisions of financial statements users²⁸. According to this principle, an examination might be carried out even though the company is no longer listed on the regulated market if the disclosure of the error is in the public interest²⁹. Conversely, material errors are not disclosed when this could put the existence of the company at risk³⁰. This occurred until December 2017 only four times³¹, as normally the interest of the financial market in knowing the error is assumed to prevail over the negative consequences for the company³². In fact, the disclosure of enforcement findings is based on the “name and shame” mechanism, which should enact a negative market reaction, discouraging companies from engaging in earnings management and prompting them to commit to high accounting quality. Additionally, enforcement announcements inform and educate other companies, potentially preventing them from committing the same error³³.

Apart from the publication of the enforcement announcement, the BaFin does not mandate the correction of the error, as this falls under the scope of the applicable accounting standards. However, it performs a follow up on the examined companies to see if errors are subsequently corrected and if any recommendations have been implemented³⁴. Additionally, the enforcement institutions refer any suspicion of criminal activity to the public prosecutor and notify the Auditor Oversight Body (previously the Chamber of German Public Auditors), if there is any indication that the auditor violated his professional duties³⁵.

²⁷ Principle 16 of CESR Standard No. 1 on financial information, available at: <https://www.esma.europa.eu/document/standard-no-1-financial-information>. Similarly, the ESMA consultation paper “Consideration of materiality in financial reporting” refrains from providing a definition of materiality and refers to the concept of materiality as defined by the IFRS. Available at: https://www.esma.europa.eu/sites/default/files/library/2015/11/2011_373_.pdf. For a discussion on materiality, see e.g. Tritschler (2013), pp. 106-108.

²⁸ IAS 8 (2018) para. 5; Zülch/Hendler (2017), pp. 533-534; Scheffler (2006), pp. 18-19.

²⁹ Sec. 342b para. 2 sentence 7 HGB.

³⁰ Sec. 109 para. 2 sentence 3 WpHG.

³¹ Thormann/Barth (2018), p. 475.

³² Gahlen/Schäfer (2006), p. 1622; Zülch (2005), p. 6; Paul (2011), p. 13.

³³ Gahlen/Schäfer (2006), p. 1623.

³⁴ E.g. FREP (2018a), p. 12.

³⁵ Sec. 342b para. 8 HGB.

5.2.2 The work of the FREP

The work of the FREP goes beyond the mere execution of examinations. In fact, the FREP works actively to prevent errors and to increase accounting quality by e.g. issuing recommendations, offering a service of pre-clearance, conducting discussions with audit firms and informing the general public³⁶. Recommendations are usually provided by the FREP after an examination is concluded and aim at improving accounting quality of future financial statements. Within the pre-clearance service, the FREP is available upon request to clarify certain critical accounting issues. This service is available for companies falling under the enforcement regime, which must provide a concrete description of the issue, together with the proposed accounting treatment and a statement of the auditor³⁷. The opinion expressed by the FREP on the matter is not binding: it is relevant only to the asking company and it can be disregarded in other enforcement examinations. Further, the FREP holds discussion with the representatives of the five biggest audit firms, and since 2018 also with the representatives of medium-sized audit firms, to increase auditors' awareness regarding the enforcement activities³⁸. Finally, the FREP takes part in technical conferences and publishes informative contents aiming at advancing the knowledge on enforcement activities and on relevant accounting issues.

A survey of PWC and DAI in June 2013 shows that companies are mostly satisfied with the work of the FREP, which they consider professionally competent³⁹. Half of the surveyed companies are convinced that the enforcement system increases the accounting quality of financial reporting in Germany. Similarly, the European Securities and Market Authority (ESMA) Peer Review indicates that the examinations are of high quality, although they should be complemented by a more informative disclosure of the findings⁴⁰. The ESMA praises the sampling model used by the FREP, citing it as an example for other enforcement institutions. Finally, the public consultations on the operations of the European supervisory authorities conducted in June 2017 by the European Commission do not see any need for further strengthening the German enforcement system⁴¹.

³⁶ E.g. FREP (2018a), p. 15.

³⁷ E.g. FREP (2015b), pp. 1-2.

³⁸ E.g. FREP (2018a), p. 23.

³⁹ The respondents of the survey are 66 companies with securities traded on the regulated market in Germany, see PWC/DAI (2013), pp. 15-21.

⁴⁰ ESMA (2017), p. 21.

⁴¹ EC (2017), p. 11.

There are several prior studies which have researched the work of the FREP and the functioning of German enforcement system. *Berger* (2010) analyses the status of enforcement in the European Union and concludes that the German enforcement institutions exercise a close control of listed companies⁴². *Tritschler* (2013) performs an evaluation of the enforcement methodology and determines that the process is objective, reliable and able to produce valid results⁴³.

Regarding accounting quality, *Ernstberger/Stich/Vogler* (2012) find that on average earnings management has decreased since the establishment of the current enforcement system in Germany⁴⁴. *Böcking/Gros/Worret* (2015) find that the German enforcement system seems to be effective in detecting earnings management, but that there is no evidence of an “educational effect” for error companies in the years following the misstatement⁴⁵. *Ebner/Hottman/Zülch* (2017) analyse if accounting quality is likely to increase for companies switching their auditor following the disclosure of enforcement findings⁴⁶. They find that an increase in quality takes place in the period gap between the error discovery and the auditor change, which leads them to conclude that in the aftermath of an enforcement announcement switching auditors does not improve accounting quality, but it is used by companies to signal higher quality.

Regarding the market reaction to the enforcement announcement, *Hitz/Ernstberger/Stich* (2012) and *Hecker/Wild* (2012) find that companies experience negative abnormal returns following the disclosure of the enforcement announcement, indicating that the market punishes infringing firms⁴⁷. *Häfele/Riediger* (2015) find that companies from the same industry as the error firms also experience negative abnormal returns around the disclosure date, which they try to compensate by enhancing the transparency of financial information. This suggests that the German enforcement system might increase the informative content of financial reporting not only among companies subject to enforcement findings⁴⁸. Finally, *Hitz/Löw/Solka* (2013) find that the audit delay, which is the period between the end of the fiscal year and the issuing of the auditor’s opinion, is

⁴² Berger (2010), p. 31.

⁴³ Tritschler (2013), pp. 112-125.

⁴⁴ Ernstberger/Stich/Vogler (2012), p. 219.

⁴⁵ Böcking/Gros/Worret (2015), p. 482.

⁴⁶ Ebner/Hottmann/Zülch (2017), p. 129.

⁴⁷ Hitz/Ernstberger/Stich (2012), p. 255; Hecker/Wild (2012), pp. 34-35.

⁴⁸ Häfele/Riediger (2015), pp. 1-3.

significantly larger for companies subject to enforcement findings, indicating that the enforcement system has also an impact on the auditing process⁴⁹.

5.3 Theoretical background and hypotheses

5.3.1 Theoretical background

According to the agency theory, two problems arise when shareholders (the principal) delegate work to the management (the agent)⁵⁰. First, shareholders and managers might have different interests and pursue different goals: e.g. investors might be more interested in increasing the long-term value of the company, while managers might be more focused on short-term achievements. Second, there is an information asymmetry between principal and agent, which together with the difference in goals might lead the agent to behave opportunistically. In this context, earnings management might occur⁵¹, which is defined as the intentional influence on financial reporting, and sometimes on the underlying transactions, to convey a certain presentation of a company's financial situation⁵².

Earnings management activities take various forms. *Dechow/Skinner* (2000) classify the accounting choices in four categories: neutral, conservative, aggressive and fraudulent⁵³. When the reported earnings mirror the company's true economic performance, one can speak of neutral accounting, whereas conservative, aggressive, and fraudulent choices imply the exercise of management discretion to deflate or inflate earnings. Conservative and aggressive accounting choices usually aim at "masking the true economic performance" of the company by managing the figures within the scope of GAAP⁵⁴. Accounting standards do not always prescribe a specific accounting treatment, leaving to managers the choice of the more accurate way to present the firm's economic situation⁵⁵. To a certain extent, earnings management might improve the informative value of financial statements⁵⁶, as it might be "a mean for managers to reveal to investors their

⁴⁹ Hitz/Löw/Solka (2013), p. 301.

⁵⁰ For an assessment and review of the agency theory, see Eisenhardt (1989), p. 58.

⁵¹ An analysis of the causes of earnings management is not the subject of this paper. For a more detailed explanation, see Healy/Wahlen (1999).

⁵² Healy/Wahlen (1999), p. 368; Wagenhofer/Ewert (2015), p. 265.

⁵³ Dechow/Skinner (2000), p. 239.

⁵⁴ Dechow/Skinner (2000), p. 240.

⁵⁵ Healy/Wahlen (1999), p. 366; Wagenhofer/Ewert (2015), p. 265.

⁵⁶ Healy/Wahlen (1999), p. 369; Ronen/Yaari (2008), p. 27; Dechow/Skinner (2000), pp. 239-240.

private expectations about the firm's future cash-flow"⁵⁷. However, the line between acceptable and undesirable earnings management is thin and depends on managers' intentions⁵⁸. In this paper, I define earnings management as the outcome of fraudulent accounting decisions, where managers violate the GAAP to convey a certain portrait of the company⁵⁹.

5.3.2 Research hypotheses

5.3.2.1 Enforcement announcements and earnings management

The existence of enforcement announcements indicates that material accounting errors have been found in financial statements. Over the years, some of the determinants of error identified by the FREP are e.g.: the complexity of IFRS, the accounting treatment of complicated transactions, the lack of knowledge of accounting standards, the financial crisis, the economic situation of the company and the conduct of earnings management⁶⁰. In fact, enforcement announcements do not provide indications on whether errors are unintentional, or the result of manipulations, failing to inform about the extent of earnings management among examined companies and about the ability of the enforcement institutions to detect manipulators.

Some studies find evidence on the relation between erroneous financial statements, identified by the US Securities and Exchange Commission (SEC) enforcement announcements, and earnings management⁶¹. However, unlike the FREP, the SEC investigates companies which display strong indications of earnings manipulation⁶². Therefore, I focus in this section on prior research specific to the German enforcement system.

The study of *Böcking/Gros/Worret* (2015) finds higher amounts of earnings management, measured with discretionary accruals, in companies with enforcement announcements compared to companies which do not have enforcement announcements⁶³. Also, their results show indications of earnings management in financial statements with enforcement findings compared to financial

⁵⁷ Beneish (2001), p. 5.

⁵⁸ Sattler (2011), p. 29.

⁵⁹ A similar definition is employed e.g. by Beneish (1999), p. 24.

⁶⁰ FREP (2018a), p. 6; FREP (2017), p. 7; FREP (2016), p. 6; FREP (2015a), p. 6; FREP (2014), p. 7; FREP (2013), p. 7; FREP (2012), p. 8; FREP (2011), p. 5; FREP (2010), appendices p. 7; FREP (2009), appendices p. 9; FREP (2008), appendices p. 6.

⁶¹ E.g. Beneish (1999), p. 24; Jones/Krishnan/Melendrez (2008), pp. 500-501.

⁶² Dechow/Ge/Larson/Sloan (2011), p. 18; Hennes/Leone/Miller (2008), p. 1488. See also the discussion in section 5.4.4 regarding the different approach of the SEC compared to that of the FREP/BaFin.

⁶³ Böcking/Gros/Worret (2015), p. 482.

statements without enforcement findings⁶⁴. *Strohmenger* (2013) finds that discretionary accruals are significantly higher for error companies in the error year compared to non-error companies, which he interprets as a sign of earnings management⁶⁵. Also, *Strohmenger* (2014) finds that error companies have lower earnings quality than the control group made up of German publicly listed firms. Again, error companies display larger amounts of discretionary accruals, which might indicate earnings management⁶⁶. *Hoehn/Strohmenger* (2013) determine that error companies are more likely to have manipulated their earnings, which they measure with accruals, than non-error companies⁶⁷. Taken together these findings might indicate that earnings management is diffused among companies subject to enforcement announcement and that the German enforcement system is efficient in findings manipulators.

Other studies focus on whether error companies display indications of engaging in earnings management. *Ernstberger/Hitz/Stich* (2012) find weak evidence that the presence of incentives for earnings management is associated with the existence of enforcement announcements, while their results do not provide support that erroneous accounting is associated with a lack of resources or of financial reporting experience⁶⁸. *Hoehn/Strohmenger* (2013) determine that the disclosure of enforcement announcements is positively associated with leverage and negatively associated with the return on assets, indicating that financial pressure and poor performance might trigger earnings management, which in turn might lead to erroneous accounting⁶⁹. Similar findings are reported by *Strohmenger* (2013), who observes that error companies are more leveraged and less profitable in the error year and over time compared to the control group⁷⁰ and by *Strohmenger* (2014), who provides evidence that in the error year (as well as in the prior and subsequent year) firms subject to enforcement findings have lower performance, are less profitable and experience higher financial pressure compared to companies without enforcement findings⁷¹. *Pasch* (2017) finds that capital market pressure and weak corporate governance are associated with the occurrence of

⁶⁴ Böcking/Gros/Worret (2015), p. 448, use non-public data on enforcement examinations, which ended without findings.

⁶⁵ *Strohmenger* (2013), p. 28.

⁶⁶ *Strohmenger* (2014), p. 296.

⁶⁷ *Hoehn/Strohmenger* (2013), p. 3.

⁶⁸ *Ernstberger/Hitz/Stich* (2012), p. 4.

⁶⁹ *Hoehn/Strohmenger* (2013), p. 22.

⁷⁰ *Strohmenger* (2013), pp. 10, 34.

⁷¹ *Strohmenger* (2014); p. 273.

enforcement findings⁷². *Ernstberger/Hitz/Stich* (2012) observe also a negative association between corporate governance and enforcement findings⁷³. Consistent with managers preferring to manage figures upwards⁷⁴, the analysis of *Toebe/Herberg/Schlüter* (2013) shows that assets and income are generally overstated in financial statements subject to enforcement announcements, while liabilities and costs are mainly understated⁷⁵. Similarly, *Hitz/Ernstberger/Stich* (2012) and *Höltken/Jana/Zülch* (2016) find some evidence that among error firms, profitability is considerably overstated, while leverage is in most cases understated⁷⁶.

Taken together these findings suggest that companies with enforcement announcements have lower accounting quality, weaker governance structure, are less profitable and more financially distressed. These factors might make them more likely to engage in earnings management, which is why misstatements among the examined companies should not be solely attributed to unintentional mistakes⁷⁷. As I expect the probability of manipulation to be higher for error companies, I formulate the following hypothesis:

H1: Financial statements subject to enforcement announcements are more likely to be manipulated than financial statements of companies without any enforcement announcements.

5.3.2.2 Error severity

Even though the errors disclosed in enforcement announcements are material⁷⁸, the impact they have on the financial statements might be quite different. Errors may affect the presentation of information in the balance sheet and the statement of comprehensive income (core statement errors) or may affect the notes, the statement of changes in equity, the statement of cash-flow or the management report (disclosure errors). Also, enforcement announcements are quite heterogeneous in terms of number of reported findings. It follows that the enforcement announcements might be classified according to the severity of the errors reported.

Prior studies provide evidence that market reactions are commensurate to the severity of the error. *Kläs/Werner* (2014) analyse market returns for clients of auditors, who audited the erroneous

⁷² Pasch (2017), p. 25.

⁷³ Ernstberger/Hitz/Stich (2012), p. 5.

⁷⁴ Defond/Jiambalvo (1991), pp. 644-646; Beneish (2001), p. 11.

⁷⁵ Toebe/Herberg/Schlüter (2013), p. 258

⁷⁶ Hitz/Ernstberger/Stich (2012), p. 266; Höltken/Jana/Zülch (2016), p. 34.

⁷⁷ A similar assumption is made by Strohmenger (2014), pp. 275-276 and Hoehn/Strohmenger (2013), p. 8.

⁷⁸ For a discussion on the concept of materiality, see section 5.2.1.

financial statements and find stronger negative market reaction when the enforcement announcement were related to the revaluation of balance sheet positions or to false or missing balance sheet items compared to missing notes/explanations⁷⁹. This might indicate that investors distinguish between errors affecting the core statements and disclosure errors. Additionally, *Klās/Werner* (2014) find that investors react negatively to announcements with multiple errors, while announcements reporting only one error are mostly ignored⁸⁰. Also, the study of *Hitz/Ernstberger/Stich* (2012) observes that severe errors are associated with stronger negative abnormal returns around the announcement date, which might indicate that investors in Germany react to the content of enforcement announcements⁸¹.

Other studies try to test whether there is an association between severe errors and the occurrence of earnings management. *Loy/Steuer* (2015), who carry out a detailed analysis of all enforcement announcements issued until 2014, find that errors affecting the core statements have an increasing effect on profit⁸², which might be an indication of earnings management. *Böcking/Gros/Worret* (2015) compare error companies with core statement errors to companies with disclosure errors. However, they do not find a significant difference in the level of earnings management between the two groups⁸³. *Strohmenger* (2013) observes that error severity is associated with high leverage⁸⁴. *Ernstberger/Hitz/Stich* (2012) find that error severity is positively associated with firm complexity and with incentives to meet earnings benchmarks, while it is negatively associated with strong corporate governance and with accounting expertise⁸⁵.

I identify severe errors with enforcement announcements reporting multiple findings where at least one pertains to the core presentation of financial statements⁸⁶. I make two assumptions here. First, I assume earnings management practices aim at influencing key financial data, which is why I expect core statement errors to be more likely associated with earnings management than disclosure

⁷⁹ Klās/Werner (2014), pp. 17-18.

⁸⁰ Klās/Werner (2014), p. 18.

⁸¹ Hitz/Ernstberger/Stich (2012), pp. 273-274.

⁸² Loy/Steuer (2015), p. 555.

⁸³ Böcking/Gros/Worret (2015), p. 481.

⁸⁴ Strohmenger (2013), pp. 32-33.

⁸⁵ Ernstberger/Hitz/Stich (2012), pp. 26-27.

⁸⁶ A similar approach, which also considers the number and the impact of the errors, can be found in Strohmenger (2013), p. 32, and Ernstberger/Hitz/Stich (2012), p. 5.

errors⁸⁷. Second, I assume that committing an error might be regarded as unintentional compared to multiple errors, which might indicate a general low accounting quality of the financial statements due to pervasive earnings management practices⁸⁸. Therefore, I expect the number of errors to be positively associated with earnings management. After classifying enforcement announcements according to the number of reported findings as well as the impact they have on the financial statements, I formulate the following hypothesis:

H2: Financial statements with severe errors are more likely to be manipulated than financial statements with non-severe errors.

5.3.2.3 Cooperation of the examined company

Enforcement examinations are usually initiated and carried out by the FREP. The company's management decides whether to collaborate with the FREP or not. If not, the examination is taken over by the BaFin, which has the authoritative power to enforce the cooperation of the company. Table 5.1 provides an overview of the examinations conducted by the BaFin until December 2017. The BaFin intervenes when companies do not voluntarily cooperate with the FREP (27.3%) or accept the results of the examination (67.9%), and when it has concrete doubts on the work conducted by the FREP (4.8%). Leaving aside this last scenario, which is a rather rare occurrence, an examination deferred to the BaFin implies a lack of cooperation of the company's management during the examination process or once the FREP communicates the results⁸⁹.

It is unclear why managers do not cooperate with the FREP but prefer undergoing the examination of the BaFin, where they might have to bear the costs of the process⁹⁰. Companies in distress, e.g. under insolvency, might fail to timely answer the requests of the FREP, thus prompting the intervention of the BaFin⁹¹. Managers might also try to "buy time" by not reacting promptly to the request of the FREP. E.g., if a company was planning a delisting or a change in market segment,

⁸⁷ A similar assumption is made by Böcking/Gros/Worret (2015), p. 481, and is also reported by Pasch (2017), p. 26. Pasch (2017), p. 26, runs a robustness check, where she drops all enforcement announcements with disclosure errors to test if the results are sensitive to the presence in the research sample of unintentional mistakes.

⁸⁸ A similar assumption is made by Böcking/Gros/Worret (2015), p. 481.

⁸⁹ A similar assumption is made by e.g. Höltnen/Jana/Zülch (2016), p. 32, Hitz/Ernstberger/Stich (2012), p. 262 and Strohmenger (2013), p. 7.

⁹⁰ According to sec. 17c of the Act Establishing the Federal Financial Supervisory Authority (Finanzdienstleistungsaufsichtsgesetz), the company does not bear the costs of the examination if the BaFin comes to a different outcome, which is in favour of the company.

⁹¹ Thormann/Zempel (2017), p. 27.

the management might refuse to cooperate with the FREP and possibly have the process dismissed by the BaFin due to the irrelevancy of the information for the public⁹². Table 5.1 shows that when companies refused to collaborate, the examination ended in 78% of the cases without enforcement findings. In contrast, in 81% of the cases where a company did not accept the outcome of the FREP's examination, the BaFin ended up confirming the findings and ordered the publication of the enforcement announcement (96%). In the remaining 19% of the cases, the BaFin either did not find any error or dismissed the case as the information was irrelevant to the capital market. In general, there might be an incentive for the management to refuse the disclosure and to wait for the BaFin to examine the matter itself, hoping either for a better outcome or at least for a delay of the announcement⁹³. Overly long examinations do not allow for the error to be corrected in the annual report following the erroneous financial statements⁹⁴. In addition to that, long delays reduce the relevancy of the announcement for investors and the negative consequences for the company. As the timely disclosure of the enforcement announcement depends on the acceptance of the results by the examined company, refuting the findings of the FREP might be a strategic decision.

Table 5.1: Overview of the BaFin examinations from July 2005 until December 2017

	No.	%	without findings	with findings	with disclosure	without disclosure
The company does not accept the results of the FREP	57	67.9%	11 (19.3%)	46 (80.7%)	44 (95.6%)	2 (4.4%)
The company does not collaborate with the FREP	23	27.3%	18 (78.3%)	5 (21.7%)	5 (100%)	0 (0%)
The BaFin has doubts on the work conducted by the FREP	4	4.8%	1 (25%)	3 (75%)	2 (66.6%)	1 (33.3%)
Total	84	100.0%	30 (35.7%)	54 (64.3%)	51 (94.4%)	3 (5.55%)

Notes: Data are collected from the BaFin annual reports for the years 2013 to 2017, which are available at: https://www.bafin.de/DE/PublikationenDaten/Jahresbericht/jahresbericht_node.html.

⁹² However, according to sec. 342b para. 2 sentence 7 HGB, if the disclosure of the error is in the public interest, an examination can continue even though the company is no longer listed on the regulated market.

⁹³ E.g. FREP (2007), p. 5; FREP (2006), p. 16.

⁹⁴ E.g. FREP (2008), p. 11.

Prior studies, which analyse the characteristics of less cooperative companies include that of *Höltken/Jana/Zülch* (2016), who find that uncooperative companies are more likely to experience an executive turnover⁹⁵, and that of *Hitz/Ernstberger/Stich* (2012), who determine that uncooperative companies are punished by negative abnormal returns around the announcement date⁹⁶. Contrary to these studies, which approximate the company's lack of cooperation with the BaFin conducting the examination, *Pasch* (2017) measures the degree of cooperativity based on non-public data of the FREP. After each examination, the FREP's auditors evaluate the cooperativity of the management: these ratings are used to build a variable, which can assume the values of 1 (good cooperation), 2 (average cooperation) and 3 (poor cooperation). *Pasch* (2017) analyses all the examinations conducted by the FREP between 2005 and 2016⁹⁷ and finds that the degree of cooperation of the company's managers, in terms of the quality of information they provide and their timely response, is negatively related to the existence of enforcement findings⁹⁸.

I assume that examinations carried out by the BaFin are an indication of the lack of cooperation of the company's managers⁹⁹, who might aim at postponing or avoiding the disclosure of the error. Companies engaging in earnings management might be more eager to avoid the disclosure, not willing to restate their numbers. Moreover, they might have more at stake (e.g. precarious financial condition, targets to reach), which is why they might try to delay the disclosure to avoid negative effects¹⁰⁰. Therefore, I expect these companies to be less inclined to cooperate with the FREP, where the lack of cooperation is assumed when the examination is taken over by the BaFin and I formulate the following hypothesis:

H3: Financial statements where the management is not cooperative are more likely to be manipulated than financial statements where the management is cooperative.

⁹⁵ Höltken/Jana/Zülch (2016), p. 37.

⁹⁶ Hitz/Ernstberger/Stich (2012), pp. 273-274.

⁹⁷ Pasch (2017), p. 15, has data, which are not publicly available, regarding examinations which ended without enforcement findings.

⁹⁸ Pasch (2017), pp. 21, 28.

⁹⁹ A similar assumption is made by e.g. Höltken/Jana/Zülch (2016), p. 32, Hitz/Ernstberger/Stich (2012), p. 262 and Strohmenger (2013), p. 7.

¹⁰⁰ However, Hecker/Wild (2012), p. 28, find that negative cumulative abnormal stock returns might increase, the longer the time-span between the balance sheet date of the erroneous financial statements and the error disclosure.

5.4 The model of Beneish (1999)

5.4.1 Measuring earnings management

In accounting literature, the level of earnings management conducted by a company is typically approximated by the amount of (discretionary) accruals¹⁰¹. Since accruals are the non-cash component of earnings, they might be easier to manipulate. The accounting literature has therefore produced various models that try to discriminate between normal accruals, which are those compatible with the company's performance, industry, and other economic characteristics, and discretionary accruals, which result from the arbitrary decisions of managers¹⁰². Even though discretionary accruals are broadly used in accounting research, they have received much criticism¹⁰³.

An alternative approach might be the use of specific accrual measures or of financial statement ratios¹⁰⁴. By analysing separately certain indicators extracted from financial statements, it might be possible to detect incentives to and consequences of misstating, as well as other characteristics that fit the profile of a manipulator. These aspects combined might provide a better insight of earnings management than aggregate accrual models¹⁰⁵. Below, I present a model which builds on specific accruals and on financial statement ratios for calculating the probability of manipulation: the model of *Beneish* (1999).

5.4.2 The m-score

Accounting professor Daniel M. Beneish introduced in 1997 and then revised in 1999 a model which uses financial statement data, to detect the probability that a company is engaging in earnings manipulation¹⁰⁶. The model structure is the following:

$$M_i = \beta' X_i + \varepsilon_i,$$

¹⁰¹ Wiedman (2002), p. 40.

¹⁰² Ronen/Yaari (2008), p. 372. In section 3.4.1, I outline several criticisms which have been raised towards discretionary accruals as a measure for audit quality. The considerations made apply in large part also to the use of discretionary accruals to measure accounting quality and earnings management.

¹⁰³ E.g. Jackson (2018), pp. 137-138, criticises the use of discretionary accrual models, which estimate the level of "normal" accruals using a cross-sectional approach, where companies are clustered by sector and year. He argues that "econometrically speaking, discretionary accruals are simply deviations from industry averages", while attributing this difference entirely to earnings management is an overconfident assumption.

¹⁰⁴ McNichols (2000), p. 335; Pasch (2017), pp. 27-28; Strohmenger (2014), p. 273.

¹⁰⁵ McNichols (2000), p. 338; Sinha (2008), p. 10; Wagner (2011), pp. 51-53.

¹⁰⁶ Beneish (1997), (1999).

M_i is a dichotomous variable that takes the value of 1 if a company is identified as a manipulator and 0 otherwise, while X_i is the matrix of the independent variables, which are designed to fit the profile of a “typical manipulator”¹⁰⁷. The idea behind it is to flag companies that show similar characteristics to known fraud cases¹⁰⁸, by picking up on distorted financial information, as well as the predisposition to engage in earnings management. *Beneish* (1999) estimates the model using a sample of 50 manipulators, matched to 1708 controls for the period 1982-1988. Manipulators are identified as being subject to an SEC enforcement release or for being exposed by the press for having committed fraud¹⁰⁹.

The model estimates the probability of manipulation based on eight variables, which are listed in Table 5.2. All variables are constructed as indices for which a higher value implies an increased likelihood of manipulation¹¹⁰. The days' sales in receivables index (DSRI) is the ratio between days' sales in receivables in the year of violation and the days' sales in receivables in the previous year. It indicates the variation of the average collection period for outstanding receivables compared to the year prior. A significant increase in the index, unless justifiable by a change in selling policies, is suspicious as it might indicate manipulated sales¹¹¹.

The gross margin index (GMI) is the ratio of the gross margin of the year t-1 to the gross margin in the year of violation t. The deterioration of the gross margin either means an increase in the cost of goods sold or a reduction in sales, which managers might want to hide by manipulating earnings.

The asset quality index (AQI) is the ratio of asset quality in year t to asset quality in t-1, where this is defined as the ratio of non-current assets other than property, plant and equipment (PP&E) to total assets. An increase in the index suggests that the company tends to defer costs by improperly capitalising assets for which future benefits are rather uncertain.

The sales growth index (SGI) is the ratio of sales in year t to sales in year t-1. An unusual growth in sales might indicate the use of phony sales to report a positive income or meet the forecasts.

¹⁰⁷ Beneish (1999), p. 26; Paolone/Magazzino (2014), p. 255.

¹⁰⁸ Beneish/Lee/Nichols (2013a), p. 57.

¹⁰⁹ To avoid including companies subject to false allegations, Beneish (1999), p. 24, requires companies to have subsequently restated their financial statements.

¹¹⁰ Beneish (1999), pp. 26-28.

¹¹¹ E.g., if managers manipulated the revenue, the corresponding receivables would not turn into cash, thus misbalancing the sales cycle.

The depreciation index (DEPI) is the ratio of the depreciation rate in t-1 to the rate in t. The depreciation rate is measured as: depreciation expenses/ (depreciation expenses + net property, plant and equipment). A value greater than 1 indicates a slowdown in the depreciation rate, which means that the company either prolonged the useful life of its assets or changed the depreciation method. Both changes, if not justified by specific circumstances, might indicate earnings manipulation.

The sales, general and administrative expenses index (SGAI) is the ratio of sales, general and administrative expenses to sales in year t-1 relative to year t. An increment of the index indicates that costs for selling have increased more than sales in the current year. If a company was to experience decreasing sales, this might motivate managers to engage in manipulations.

The leverage index (LVGI) is the ratio of total debt to total assets in year t relatively to year t-1. Increasing financial pressure might be an incentive to manipulate, e.g. to avoid higher borrowing costs or breaking debt covenants.

Total accruals to total assets (TATA) is the ratio of net income before extraordinary items minus cash-flow from operations to total assets¹¹². Higher accruals indicate that earnings are less cash-based and might signalise the existence of earnings management.

The variables aim at determining the likelihood of manipulation by capturing distortions, which might be related to earnings management¹¹³, e.g. the misbalanced relation between receivables and sales (DSRI), the unusual capitalisation of expenses (AQI), any non-justifiable changes in depreciation (DEPI), sales growing too quickly (SGI), and high levels of non-cash-based earnings (TATA) Additionally, the model accounts for several incentives for earnings management¹¹⁴, e.g. a deteriorating gross margin (GMI), an over proportional increase in sales, general and administrative expenses (SGAI), and increased leverage (LEVI). To measure the probability of manipulation, the variables are combined in the following formula:

$$M - Score = -4.84 + 0.920 * DSRI + 0.528 * GMI + 0.404 * AQI + 0.892 * SGI + 0.115 * DEPI - 0.172 * SGAI + 4.679 * TATA - 0.327 * LVGI$$

¹¹² In Beneish (1999) the variable is measured according to the balance sheet approach. See the notes to Table 5.2.

¹¹³ Beneish/Nichols (2007), p. 6.

¹¹⁴ Beneish/Nichols (2007), p. 6.

Table 5.2: Variables

$$DSRI = \frac{receivables_{i,t}/sales_{i,t}}{receivables_{i,t-1}/sales_{i,t-1}}$$

$$GMI = \frac{(sales_{i,t-1} - costofgoodssold_{i,t-1})/sales_{i,t-1}}{(sales_{i,t} - costofgoodssold_{i,t})/sales_{i,t}}$$

$$AQI = \frac{(1 - (currentassets_{i,t} + PPE_{i,t})/totalassets_{i,t})}{(1 - (currentassets_{i,t-1} + PPE_{i,t-1})/totalassets_{i,t-1})}$$

$$SGI = \frac{sales_{i,t}}{sales_{i,t-1}}$$

$$DEPI = \frac{Depreciation_{i,t-1}/(Depreciation_{i,t-1} + PPE_{i,t-1})}{Depreciation_{i,t}/(Depreciation_{i,t} + PPE_{i,t})}$$

$$SGAI = \frac{sgaexpense_{i,t}/sales_{i,t}}{sgaexpense_{i,t-1}/sales_{i,t-1}}$$

$$LVGI = \frac{(LTD_{i,t} + currentliabilities_{i,t})/totalassets_{i,t}}{(LTD_{i,t-1} + currentliabilities_{i,t-1})/totalassets_{i,t-1}}$$

$$TATA = \frac{netincomebeforeextraordinaryitems_{i,t} - CFO_{i,t}}{totalassets_{i,t}}$$

Notes: The index TATA was originally measured by Beneish (1999) following the so-called balance sheet approach (see formula below). More recently, Beneish/Lee/Nichols (2013), p. 60, have suggested the use of the direct method, where cash-flow from operations is subtracted from net income before extraordinary items and divided by total assets.

$$TATA = \frac{\begin{pmatrix} \Delta Current\ assets - \Delta Cash - \\ \Delta Current\ liabilities - \Delta Current\ maturities\ of\ LTD \\ - \Delta Income\ tax\ payable \\ - Depreciation\ and\ Amortization \end{pmatrix}}{Total\ assets}$$

Beneish (1999) estimates the coefficients with the help of a probit regression. The model can make two types of errors: classifying a company as non-manipulator even though it engages in earnings management (type I) and classifying a company as a manipulator when it is not (type II). Type I and type II errors are associated with different costs. Beneish (1999) argues that in the quarter when the manipulation becomes known to the public, manipulators lose around 40% of their market value on a risk-adjusted basis. Conversely, in the absence of fraud, their investment would have likely

had over the same period an increase in value between 1% and 2%¹¹⁵. Therefore, he estimates the costs for type I errors to be 20 to 40 times higher than costs for type II errors¹¹⁶. To minimise the costs of misclassification, *Beneish* (1999) identifies a cut-off value of -1.78¹¹⁷. A company is then classified as a manipulator in a given year, if the m-score exceeds the value of -1.78. By using this threshold, the model is able to correctly classify 74% of the manipulators and 86.2% of non-manipulators in the original sample¹¹⁸.

5.4.3 Prior studies using the m-score

There are a fair number of studies, which have used the m-score to detect earnings management. The m-score has become popular as it is easy to calculate and can be employed not only by scholars, but also by investors, auditors and researchers to assess the likelihood of manipulation for companies with available financial data¹¹⁹. However, two issues arise when comparing previous studies employing the m-score. The original m-score formula has 8 variables, but some studies employ an alternative version of the model with only 5 variables, which appeared in a 1994 working paper of *Beneish* (no longer available). This 5-variables version of the model has not been empirically tested out of sample and recently the author has advised against using it¹²⁰. Another controversial point is the cut-off value for distinguishing between manipulators and non-manipulators. *Beneish* (1999) suggests the use of -1.78 as a threshold¹²¹, but other authors have employed different cut-off values¹²². This heterogeneity in the number of variables and in the choice of cut-off hinders the comparison of the results of prior research.

In this section, I merely present a selection of studies, which have employed the model of *Beneish*. Some studies test the accuracy of the model by applying it to known fraud cases. *Grove/Cook* (2004) measure 5 of the 8 variables of the model (DSRI, GMI, AQI, SGI and TATA) for four high-

¹¹⁵ *Beneish/Nichols* (2007), p. 1, argue that by incorrectly identifying a non-manipulator as a manipulator, an investor loses the expected return on that investment. However, as the investor might be able to invest in other stocks with comparable returns, the costs related to a type II error are minimal.

¹¹⁶ Conversely, *Grove/Basilico* (2008), p. 14, argue that type I errors are 40 to 60 times costlier than type II errors.

¹¹⁷ *Beneish* (1999), pp. 32-33.

¹¹⁸ *Beneish* (1999), pp. 32-33, tests this cut-off value on a hold-out sample, which includes 24 manipulators and 624 controls during 1989-1992. In the hold-out sample, the model misclassifies 50% of manipulators and 7.2% of non-manipulators.

¹¹⁹ *Grove/Cook* (2004), p. 144.

¹²⁰ In an e-mail exchange in the month of August 2017.

¹²¹ *Beneish* (1999), pp. 32-33; *Beneish/Lee/Nichols* (2013a), p. 60. See also the considerations made in 5.4.2.

¹²² E.g. *Repousis* (2016), p. 1068; *Ahn/Lihn* (2016), p. 19; *Grove/Basilico* (2008), p. 14.

profile fraud cases: Enron, WorldCom, Global Crossing and Qwest¹²³. They report that only the variables GMI and SGI indicate the presence of earnings management, but they do not calculate the m-score. *Roxas* (2011) calculates the m-score for a sample of companies with known revenue recognition violations and shows how in most cases the m-score can flag manipulators in the year the fraud was discovered and one year in advance¹²⁴. *MacCarthy* (2017) finds that the m-score classifies Enron as a manipulator in each of the three years preceding the discovery of the fraud in 2001¹²⁵. Similarly, *Beneish/Nichols* (2007) find that 12 out of 20 high profile fraud cases from the US during 1998-2002 are identified by the m-score a year and half before the fraud is made public¹²⁶. *Mehtaa/Bhavani* (2017) report that the m-score fails to classify Toshiba as a manipulator in the years 2008 to 2014, which is the period when the company committed fraud¹²⁷. *Kamal/Salleh/Ahmad* (2016) find that the m-score is successful in detecting 14 out of 17 Malaysian companies prosecuted for fraudulent financial reporting¹²⁸. *Tarjo* (2015) finds that the m-score correctly classifies 77% of manipulators and 80% of non-manipulators in a sample of 35 Indonesian companies committing fraud over the period between 2001 and 2014 and 35 industry and size-matched controls¹²⁹.

As there is evidence that the m-score is associated with the magnitude of the fraud¹³⁰, some authors use it as a proxy for fraud or misstatement risk¹³¹. Other studies employ it to determine the rate of manipulators in a given sample. For example, *Paolone/Magazzino* (2014) calculate the m-score of a sample of Italian stock corporations from the textile, food, clothing, automotive and metallurgic industry¹³². Their results show that 51.4% of the companies have a high probability of manipulating earnings over the period 2005-2012. *Corsi/DiBernardino/DiCimbrini* (2015) find for a group of Italian academic spin-offs during the period 2009-2010 that 33% of the companies might be

¹²³ Grove/Cook (2004), pp. 133, 144.

¹²⁴ Roxas (2011), p. 64.

¹²⁵ MacCarthy (2017), p. 164.

¹²⁶ Beneish/Nichols (2007), p. 2.

¹²⁷ Mehta/Bhavani (2017), p. 701.

¹²⁸ Kamal/Salleh/Ahmad (2016), p. 24.

¹²⁹ Tarjo (2015), p. 928.

¹³⁰ Jones/Krishnan/Melendrez (2008), p. 500.

¹³¹ E.g. Cassell/Myers/Myers/Seidel (2016), p. 2; Sihna (2008), p. 2; DeFond/Lim/Zang (2016), p. 81.

DeFond/Lim/Zang (2016), use the average annual decile rank of the m-score to estimate the expected misstatement risk. Further, there is evidence that the m-score is useful in predicting future stock returns, see e.g. Beneish/Nichols (2007), p. 2; Beneish/Nichols/Lee (2011), p. 2; Beneish/Lee/Nichols (2013a), p. 75.

¹³² Paolone/Magazzino (2014), pp. 256-257.

manipulators¹³³. *Franceschetti/Koschtial* (2013) calculate the m-score for a group of bankrupt and non-bankrupt Italian companies¹³⁴. Their results lead to reject the hypothesis that companies facing the risk of insolvency are more likely to resort to earnings manipulation. *Repousis* (2016) finds that 33% of non-financial Greek companies (Société Anonyme and Limited Liabilities companies) are flagged as manipulators by the Beneish model¹³⁵. A recent study by *Ahn/Linh* (2016) of the Vietnamese market reports that the m-score classifies 48.4% of non-financial listed companies as manipulators¹³⁶.

Finally, some authors try to improve the original model. *Dechow/Ge/Larson/Sloan* (2011), building on the work of *Beneish* (1999), develop the F-score, where additional non-financial and market-related variables are added to the model¹³⁷. *Feruleva/Shtefan* (2017) modify the original model to fit the Russian market¹³⁸. *Grove/Basilico* (2008) exclude the variables, which are not significant in the original model (DEPI, SGAI and LVGI) and call the 5-variables model z-score¹³⁹. By using a cut-off value of -1.99 , they report that the z-score correctly classifies 76% of manipulators one year before the fraud becomes known to the public and 66% of manipulators two years before the scandal¹⁴⁰.

5.4.4 Brief discussion on the need to re-estimate the model of Beneish (1999)

Previous studies, including this one, apply the model of *Beneish* (1999) with its original coefficients. The sample used to estimate and test the coefficients include 74 U.S. companies that manipulated their earnings from 1982 to 1992 and 2332 non-manipulators matched by industry and year. The question arises whether the model should be updated and re-estimated. In this section, I briefly discuss some points, which need to be considered before updating and re-estimating the model.

First, the original model strongly relies on accounting theory to identify the typical traits of earnings manipulators. An attentive analysis of the more recent literature on fraud is necessary to assess if

¹³³ Corsi/Di Bernardino/Di Cimbrini (2015), pp. 75-76.

¹³⁴ Franceschetti/Koschtial (2013), p. 2.

¹³⁵ Repousis (2016), p. 1072.

¹³⁶ Ahn/Linh (2016), p. 21.

¹³⁷ Dechow/Ge/Larson/Sloan (2011), pp. 21-22.

¹³⁸ Feruleva/Shtefan (2017), p. 33.

¹³⁹ Grove/Basilico (2008), pp. 13-14, 31-32.

¹⁴⁰ Grove/Basilico (2008), p. 29.

the model variables are still relevant or if other aspects should be taken into consideration. Special attention should be dedicated to country-specific aspects which are likely to affect the manipulation of earnings. E.g. *Feruleva/Shtefan* (2017) modify the variable AQI arguing that under Russian GAAP certain items reported under “other current assets” do not qualify as proper assets and add two new variables to account for Russian-specific accounting treatments¹⁴¹. Second, the popularity of the model is due to the fact that it is accessible to all users of financial statements, which is why any updated version of the model should still be measurable based on financial statements data. Third, another aspect to consider when re-estimating the model is the necessity to gather together an appropriate sample of companies that have committed fraud. *Beneish/Lee/Nichols* (2013b) state that a re-estimation of the model with current data might make sense¹⁴². However, they caution that, as the model is highly sensitive to the composition of the manipulator sample, it should include only fraudulent companies. The presence of non-manipulators in the manipulator sample is likely to weaken the predictive power of the model, which might incur a higher number of type II errors (classifying a company as a manipulator when it is not). *Beneish* (1999) argues that the cost of type II errors is minimal for investors as he assumes, they can choose an alternative investment with a similar expected return. However, the cost of type II errors might be significant for auditors, as audit costs increase due to the identified higher risk of fraud, and for enforcement institutions if they initiate an examination based on the (false) indication of earnings management.

The original estimation sample includes companies subject to enforcement actions by the SEC¹⁴³. Unlike the FREP, the SEC, due to the limited budget, does not randomly sample the companies to be examined, but it selects those with strong indications of fraud. A company admitting some wrongdoing (e.g. restatement), indications from the press, and whistle-blowers are usually the trigger for an SEC examination¹⁴⁴. It follows that being subjected to an SEC enforcement announcement is a strong indication of fraud¹⁴⁵. Conversely, the examinations of the FREP are only partially initiated due to indications of misstatements and might be related to a lack of

¹⁴¹ *Feruleva/Shtefan* (2017), p. 37. The two variables are the income quality index (IQI), which is the change in the ratio between other income and sales, and the dummy variable ACC, which is set to 1 if the company chooses to charge administrative costs to the core product expenses.

¹⁴² *Beneish/Lee/Nichols* (2013b), p. 14.

¹⁴³ Some companies are identified according to the indication in the press of having committed fraud, see section 5.4.2.

¹⁴⁴ *Dechow/Ge/Larson/Sloan* (2011), p. 18; *Hennes/Leone/Miller* (2008), p. 1488.

¹⁴⁵ *Dechow/Sloan/Sweeney* (1996), p. 2, argue that the inclusion of spectacular fraud cases in the sample on the one hand increases the test power, but on the other hand it limits the generalisability of the results.

resources or expertise, as well as the complexity of IFRS¹⁴⁶. For these reasons, I avoid updating the model in this study. Researchers willing to update the model should make sure they recognise and include companies in the sample which have manipulated their earnings.

5.5 Research design

5.5.1 Sample selection

Companies with accounting errors are identified based on the enforcement announcements published in the Federal Gazette (Bundesanzeiger)¹⁴⁷. From the initial sample of 258 enforcement announcements¹⁴⁸, double announcements or amendments of already released announcements are eliminated. To have a more homogenous sample in terms of accounting data, I exclude announcements regarding half-year financial statements, foreign companies, as well as single financial statements and consolidated financial statements not reporting under IFRS. Observations regarding financial companies (banks, insurance, real estate) are also dropped as the model of *Beneish* (1999) is explicitly not designed for financial companies¹⁴⁹. Furthermore, 12 announcements are dropped due to the lack of data in Thomson Reuters Eikon. Since the formula for the m-score requires data for the previous year, one set of financial statements is excluded as the company applied IFRS for the first time in the error year. This leaves a dataset of 121 announcements regarding financial years ending between 2004 and 2015. Table 5.3 provides an overview of the sample selection.

¹⁴⁶ For a comparison between SEC and FREP enforcement announcements, see e.g. Hitz/Ernstberger/Stich (2012), pp. 269-270.

¹⁴⁷ The initial sample is the same as described in Fiallo/Hecker (2019a), p. 228 Table 2, as well as in sections 3.4.2 and 4.3.1. However, data have been collected, independently from the studies mentioned above, from the company's consolidated financial statements and from Thomson Reuters Eikon.

¹⁴⁸ The search on the website of the Federal Gazette (<https://www.bundesanzeiger.de>) for the chosen period produces 250 results. As explained in the note to Table 2 in Fiallo/Hecker (2019a), p. 228, eight enforcement announcements regarding different sets of financial statements (e.g. yearly and half-yearly) are here separately accounted for.

¹⁴⁹ Beneish/Lee/Nichols (2013a), p. 60.

Table 5.3: Sample selection - Error group

Enforcement announcements issued in the Federal Gazette until 31/12/2017	258
Announcements disclosed twice or amended	- 17
Announcements regarding half-year financial statements	- 30
Announcements regarding foreign companies	- 19
Announcements regarding annual financial statements (local GAAP)	- 11
Announcements regarding consolidated financial statements (US-GAAP or local GAAP)	- 8
Announcements regarding financial companies	- 39
Announcements regarding companies for which data were not available	- 12
Announcements regarding financial statements applying IFRS for the first time	- 1
Final sample	121

Notes: This table is based on Fiallo/Hecker (2019a), p. 228 Table 2.

In order to analyse the relationship between enforcement announcements and earnings management, I build a control group with companies subject to the enforcement regime, but which have not disclosed any enforcement announcements as of 31 December 2017. The control group is drawn from the list of all companies falling under the enforcement regime as of 1 July 2017¹⁵⁰. Of the initial 561 companies falling under the enforcement regime, 112 companies are excluded as they issued at least one enforcement announcement before 31 December 2017¹⁵¹. Following the same selection process for the error group, foreign companies (22) and companies from the financial sector are eliminated (60). Another 73 companies are unavailable in Thomson Reuters Eikon. This leaves a dataset of 294 companies with financial data for the period between 2003 and 2015¹⁵². After eliminating firm/year observations regarding financial statements not prepared under IFRS and for which data are not available, there remains a dataset of 2440 firm-year observations.

¹⁵⁰ The current list is publicly available at:

https://www.bafin.de/SharedDocs/Downloads/DE/Liste/dl_li_fis_enforcement.pdf;jsessionid=529B459F07AB1B8ACF16D1B3D42EC540.2_cid381?__blob=publicationFile&v=6. Lists for prior years can be requested from the BaFin.

¹⁵¹ I caution that the control group might include companies which have not been examined as of 31 December 2017. Also, there might be companies, whose financial statements were found erroneous during the enforcement examination, but which did not publish an enforcement announcement, as either the information was irrelevant to the financial market or because the disclosure could critically damage the company's interest (sec. 109 para. 2 WpHG). Until 31 December 2017 there are 8 such cases. Besides the 3 cases reported in Table 5.1, there are 5 more cases where the company accepted the findings of the FREP, but they were not disclosed in an enforcement announcement, see BaFin (2013), p. 181. As reported by Böcking/Gros/Worret (2015), p. 448, and Pasch (2017), p. 6, a less noisy control group would be made by firm/year observations where the enforcement examination does not end with enforcement findings. However, these data are not publicly available, and my request to the BaFin to obtain them for research purposes has been denied, see also Strohmenger (2014), p. 299.

¹⁵² Data for the year 2003 are necessary to measure the m-score for the earliest error year, which is 2004.

Table 5.4: Sample selection - Non-error group.

Companies subject to enforcement as of 1/7/2017	561
Companies with enforcement announcements before 31/12/2017	- 112
Foreign companies	- 22
Financial companies (banks, insurance, real estate)	- 60
Companies not available in Thomson Reuters Eikon	- 73
Final sample	294

Notes: The 294 companies are distributed among the following sectors: consumer discretionary (23.5%), telecommunication services (1.4%), materials (8.5%), information technologies (18%), consumer staples (3.1%), health care (11.2%), industrials (30.6%) and utilities (3.7%).

Tables 5.5 and 5.6 report the sector distribution, according to the Global Industry Classification Standard (GICS), and the year distribution in the sample. Table 5.7 compares size, profitability, and leverage data of companies in the error and non-error group. Error companies are on average smaller, both in terms of total assets and volume of sales, compared to non-error companies. Also, they are less profitable, as shown by RoA and RoE, than non-error companies. The two measures of leverage indicate that error companies are less liquid and more indebted than non-error companies. The statistics for the error group in Table 5.7 refer to the error year. The results for the prior year, which are not reported, are qualitative similar. These results are consistent with those of *Beneish (1999)*, who finds that manipulators in the year prior to committing fraud are smaller, quickly growing, less profitable and more leveraged than non-manipulators¹⁵³.

Table 5.5: Sector distribution

Sector	Error group		Non-error group	
	Observations	Percentage	Observations	Percentage
Consumer discretionary	32	26.4%	521	21.4%
Telecommunication services	5	4.1%	30	1.2%
Materials	6	5.0%	171	7.0%
Information technologies	32	26.4%	516	21.1%
Consumer staples	5	4.1%	91	3.7%
Health care	11	9.1%	293	12.0%
Industrials	27	22.3%	772	31.6%
Utilities	3	2.5%	46	1.9%
Total	121	100.0%	2.440	100.0%

Notes: Data are retrieved from Thomson Reuters Eikon. The relevant data item is TR.GICSsector.

¹⁵³ Beneish (1999), p. 25.

Table 5.6: Year distribution

Year	Error group		Non-error group	
	Observations	Percentage	Observations	Percentage
2004	2	1.7%	112	4.6%
2005	23	19.0%	178	7.3%
2006	17	14.0%	192	7.9%
2007	11	9.1%	208	8.5%
2008	14	11.6%	209	8.6%
2009	14	11.6%	210	8.6%
2010	4	3.3%	214	8.8%
2011	6	5.0%	216	8.9%
2012	12	9.9%	220	9.0%
2013	4	3.3%	224	9.2%
2014	8	6.6%	229	9.4%
2015	6	5.0%	228	9.3%
Total	121	100.0%	2.440	100.0%

Table 5.7: Comparison in terms of size, profitability and leverage in the error year

Variables	Error group			Non-error group			Wilcoxon Test	Median test
	Mean	Median	Obs.	Mean	Median	Obs.	p-value	p-value
<i>Size</i>								
Total assets	1113957582	131026000	121	3369438994	238006180	2440	0.000	0.004
Sales	1186382048	128853000	121	2890627470	289166000	2440	0.000	0.000
<i>Profitability</i>								
Net income	11447765	885000	121	122607530	8341625	2440	0.000	0.000
RoE	-0.303	0.061	121	0.102	0.103	2440	0.001	0.000
RoA	-0.044	0.012	121	0.019	0.040	2440	0.000	0.000
<i>Leverage</i>								
cfo/total assets	0.001	0.047	121	0.067	0.080	2440	0.000	0.000
total liabilities/ total assets	0.671	0.646	121	0.568	0.569	2440	0.000	0.001

Notes: I run a Wilcoxon rank-sum test and a non-parametric equality-of-medians test. The p-values indicate at which level it is possible to reject the null hypothesis that the two samples were drawn from a population with the same median.

5.5.2 Methodology

To test hypothesis H1, whether financial statements subject to enforcement announcements are more likely to be manipulated than financial statements of companies without enforcement announcements, I run some univariate tests. First, I calculate the m-score for each firm/year observation in the sample and present the results according to year and sector. A company is classified as a manipulator when the m-score is bigger than -1.78, otherwise as a non-

manipulator¹⁵⁴. Second, I calculate the mean and median of each financial statement ratio used to estimate the m-score and run the Wilcoxon rank-sum test and the equality-of-medians test to verify whether there are significant differences between the error and non-error group¹⁵⁵. Third, I build the variable ERROR, which is a dichotomous variable, assuming the value of one, if the financial statements of a company are subject to an enforcement announcement, and zero otherwise. Then I run several univariate probit regressions in which each financial statement ratio is used as a predictor of the existence of enforcement announcements.

To test hypothesis H2 and H3 I run the same univariate tests, this time only on the sample of error companies. Hypothesis H2 states that financial statements with severe errors are more likely to be manipulated than financial statements with non-severe errors. Error severity is approximated with the variable ERRORSEV, which assumes the value of 1 when the company has multiple errors with at least one pertaining to the core presentation of financial statements (balance sheet and income statement)¹⁵⁶. In the error group, I identify 73 enforcement announcements with severe errors¹⁵⁷.

Hypothesis H3 states that financial statements prepared by uncooperative management are more likely to be manipulated than financial statements where managers cooperate with the FREP. The lack of cooperation is assumed when the BaFin takes over the examination. The variable ERRORINST assumes the value of 1 if the enforcement examination is carried out by the BaFin, and zero otherwise¹⁵⁸. In the sample 26 examinations are conducted by the BaFin and 95 by the FREP.

Finally, all three hypotheses are tested in a multivariate analysis. With the help of a probit regression, I analyse if the eight ratios proposed by Beneish are useful in predicting alternatively the existence of enforcement announcements, the severity of the error and the lack of cooperation of a company (ERROR, ERRORSEV, ERRORINST). *Beneish* (1999) suggests the use of either an

¹⁵⁴ For a discussion on the cut-off value, see section 5.4.2.

¹⁵⁵ Similar tests are run by Beneish (1999), p 25.

¹⁵⁶ Data on the number of errors are collected from the enforcement announcements. Data on the impact of the error are collected from the enforcement announcement, the company's annual reports for subsequent years and when available from restatements.

¹⁵⁷ Among the 121 observations in the error group, 96 have core statement errors and 85 have multiple errors.

¹⁵⁸ This information is obtained from the enforcement announcement.

“unweighted” probit or of a weighted exogenous sample maximum likelihood (WESML) probit¹⁵⁹. The WESML has the advantage of limiting the bias resulting from the overrepresentation of manipulators in the sample compared to the population, but it does not provide qualitatively different results¹⁶⁰. It requires, however, the ratio of manipulators in the population, which is unknown. As both methods are likely to yield similar results, I prefer using the unweighted probit to estimate the coefficients¹⁶¹.

The model is the following¹⁶²:

$$\begin{aligned}
 &ERROR_{i,t} \text{ or } ERRORSEV_{i,t} \text{ or } ERRORINST_{i,t} \\
 &= \alpha + \beta_1 * DSRI_{i,t} + \beta_2 * GMI_{i,t} + \beta_3 * AQI_{i,t} + \beta_4 * SGI_{i,t} + \beta_5 * DEPI_{i,t} + \beta_6 \\
 &* SGAI_{i,t} + \beta_7 * LVGI_{i,t} + \beta_8 * TATA_{i,t} + \sum_{j=1}^{N-1} \beta_j * sector_{i,j} + \sum_{k=1}^{N-1} \beta_k * year_{i,k} \\
 &+ \varepsilon_{i,t}
 \end{aligned}$$

All independent variables are explained in Table 5.2. As the level of sales, general and administrative expenses is not separately reported when the income statement is classified according to the nature of expenses, the necessary data for calculating SGAI are missing for 814 firm/year observations¹⁶³. In these cases, I follow the suggestion of *Beneish* (1999) and set the value of SGAI equal to one, which is the neutral value for this index¹⁶⁴.

5.6 Results

5.6.1 Univariate results: enforcement announcements and earnings management

To test Hypothesis H1, whether financial statements subject to enforcement announcements are more likely to be manipulated than financial statements of companies with no enforcement announcements, I run three tests. The first test involves calculating the m-score for each firm/year observation in the sample. Table 5.8 shows that 33% of the financial statements in the error group

¹⁵⁹ Beneish (1999), p. 26.

¹⁶⁰ Zmijewski (1984), pp. 77-80.

¹⁶¹ Beneish (1999), p. 28. It should be noted that the coefficients used to calculate the m-score are also estimated with an unweighted probit.

¹⁶² As the sample contains unbalanced panel data, I add year and sector fixed-effects to the model. All continuous variables are winsorized as the 1% and 99% to minimise the effect of outliers as well as of small denominators.

¹⁶³ Thereof 56 observations in the error group and 758 in the non-error group.

¹⁶⁴ Beneish (1999), p. 28.

are flagged as manipulated against 18.6% in the non-error group, which might indicate that earnings management is more common in the error group than in the non-error group. When looking at the year distribution, there is no visible trend in the error group, while among non-error companies the proportion of manipulators seems to have decreased in more recent years. In the error group the percentage of manipulated financial statements seems to be extremely high in the materials and consumer staples sectors ($> 60\%$)¹⁶⁵, while in the non-error group the proportion does not exceed 23% in any sector. Conversely, the consumer staples sector in the non-error group has a particularly low incidence of manipulated financial statements (5%). In the error group, consumer discretionary is the sector with the lowest percentage of manipulated financial statements (13%).

The second test involves calculating the mean and median of each ratio in the error and non-error group. Like *Beneish* (1999), I run two non-parametric tests: the Wilcoxon rank-sum test and the equality-of-medians test, to verify if the variables are significantly different between the two groups¹⁶⁶. Table 5.9 shows that error and non-error companies differ according to the depreciation index (p-values < 0.05) and the leverage index (p-values < 0.01). One of the assumptions made is that manipulation might take place through a slow-down in the depreciation rate aiming at increasing the reported profit. However, the results show that the depreciation index (DEPI) is smaller for the error firms (0.954) than for non-error firms (1.002). While the median value for the non-error group is close to the neutral value of 1, which indicates no change compared to the previous year, a value smaller than 1 means an acceleration of the depreciation rate in the current year. The leverage index (LVGI) is higher for error companies (1.025) than for the controls (0.99), indicating an increase in financial pressure compared to the previous year, which might be an incentive to engage in earnings management. This finding is in line with prior studies which also find enforcement announcements to be associated with leverage¹⁶⁷. There is limited evidence that the variables DSRI and SGAI differ in the two groups¹⁶⁸, which is why I refrain from interpreting these results. The m-score is higher for the error group, which might indicate a higher likelihood

¹⁶⁵ The significance of this result is limited by the number of observations in the materials and consumer staples sectors, which are respectively 6 and 5.

¹⁶⁶ Beneish (1999), p 25.

¹⁶⁷ Strohmenger (2013), pp. 10, 34; Hoehn/Strohmenger (2013), p. 22; Strohmenger (2014); p. 273; Pasch (2017), p. 25.

¹⁶⁸ According to the Wilcoxon test, they are significantly different respectively at the 5% and 10% level. According to the median test, they are both not significant.

of manipulation, but it is significant only according to the Wilcoxon test (p-value < 0.05). Finally, there is no evidence that the variables GMI, AQI, SGI and TATA are significantly different in the error and non-error group.

Table 5.8: Classification according to the m-score per year and sector - H1

	ERROR = 1				ERROR = 0			
	Manipulators		Non-manipulators		Manipulators		Non-manipulators	
2004	0	0%	2	100%	23	21%	89	79%
2005	6	26%	17	74%	51	29%	127	71%
2006	9	53%	8	47%	54	28%	138	72%
2007	5	45%	6	55%	59	28%	149	72%
2008	5	36%	9	64%	36	17%	173	83%
2009	5	36%	9	64%	29	14%	181	86%
2010	1	25%	3	75%	42	20%	172	80%
2011	1	17%	5	83%	34	16%	182	84%
2012	3	25%	9	75%	28	13%	192	87%
2013	1	25%	3	75%	33	15%	191	85%
2014	1	13%	7	87%	36	16%	193	84%
2015	3	50%	3	50%	29	13%	199	87%
	40	33%	81	67%	454	19%	1986	81%

	ERROR = 1				ERROR = 0			
	Manipulators		Non-manipulators		Manipulators		Non-manipulators	
Consumer Discretionary	4	13%	28	87%	101	19%	420	81%
Telecom. Services	2	40%	3	60%	6	20%	24	80%
Materials	4	67%	2	33%	19	11%	152	89%
Information Technology	15	47%	17	53%	119	23%	397	77%
Consumer Staples	3	60%	2	40%	5	5%	86	95%
Health Care	2	18%	9	82%	66	23%	227	77%
Industrials	9	33%	18	67%	133	17%	639	83%
Utilities	1	33%	2	67%	5	11%	41	89%
	40	33%	81	67%	454	19%	1986	81%

Notes: The variable ERROR is coded 1 when the financial statements are subject to an enforcement announcement, and zero otherwise. The probability of manipulation is measured with the following equation $m\text{score}_{i,t} = -4.84 + 0.920 * DSRI_{i,t} + 0.528 * GMI_{i,t} + 0.404 * AQI_{i,t} + 0.892 * SGI_{i,t} + 0.115 * DEPI_{i,t} - 0.172 * SGAI_{i,t} + 4.679 * TATA_{i,t} - 0.327 * LVGI_{i,t}$. Financial statements are classified as manipulated when the m-score is bigger than -1.78.

Table 5.9: Descriptive statistics - H1

Variables	ERROR = 1			ERROR = 0			Wilcoxon test	Median test
	Mean	Median	Obs.	Mean	Median	Obs.	p-value	p-value
DSRI	1.1929	1.0501	121	1.0522	0.9994	2440	(0.049)	(0.304)
GMI	0.9713	0.9996	121	1.0105	0.9985	2440	(0.625)	(0.512)
AQI	1.2412	0.9889	121	1.0954	0.9901	2440	(0.645)	(0.929)
SGI	1.1401	1.0762	121	1.1016	1.0625	2440	(0.489)	(0.512)
DEPI	1.0315	0.9535	121	1.0417	1.0015	2440	(0.010)	(0.033)
SGAI	1.1304	1.0000	121	1.0558	1.0000	2440	(0.094)	(0.584)
LVGI	1.0817	1.0248	121	1.0173	0.9897	2440	(0.000)	(0.004)
TATA	-0.0492	-0.0343	121	-0.0468	-0.0402	2440	(0.936)	(0.638)
m-score	-1.8608	-2.1015	121	-2.1414	-2.2159	2440	(0.046)	(0.112)

Notes: The variable ERROR is coded 1 when the financial statements are subject to an enforcement announcement and zero otherwise. I run a Wilcoxon rank-sum test and a non-parametric equality-of-medians test. The p-values indicate at which level it is possible to reject the null hypothesis that the two samples are drawn from a population with the same median.

The third test involves the running of several univariate probit regressions in which each variable is used as predictor of the existence of enforcement announcements. Table 5.10 shows that all models have limited explanatory power, but that according to the X^2 test they are all statistically significant (p-values <0.01). Five variables are significantly positively associated with the existence of enforcement announcements: DSRI, AQI, SGAI, LVGI and the m-score. The results show that the probability of having enforcement announcements increases for companies with longer days' sales in receivables compared to the previous year, with a higher capitalisation of non-current assets other than PP&E, with a disproportionate increase in the sales, general and administrative expenses (weak significance) and which experience increasing financial pressure. Also, the m-score is positively associated with enforcement announcements, indicating that error companies might be more likely to manipulate their earnings.

In conclusion, the univariate results show that error companies have higher m-scores than non-error companies and that there is a positive association between the m-score and enforcement announcements. These findings might indicate that earnings management is more common in the error group than in the non-error group. Further, I observe that error companies are more leveraged than non-error companies. This is in line with prior studies and might suggest that error companies might have a motive for engaging in earnings management. Another motive might be a decrease in sales, although the sales, general and administrative expenses index is only weakly significant. Further, the results provide some indications that earnings management might take place through

the manipulation of sales and the improper capitalisation of expenses, while there is no evidence that it might be conducted by adjusting the depreciation of assets.

Table 5.10: Univariate regressions - HI

Variables	Coeff.	Constant	Prob > X^2	Pseudo R ²	Obs.
DSRI	0.258*** (0.001)	-1.956*** (0.000)	0.010	0.095	2561
GMI	-0.493 (0.984)	-1.182*** (0.000)	0.005	0.048	2561
AQI	0.154*** (0.006)	-1.849*** (0.000)	0.006	0.059	2561
SGI	0.142 (0.114)	-1.831*** (0.000)	0.001	0.014	2561
DEPI	-0.037 (0.621)	-1.634*** (0.000)	0.000	0.001	2561
SGAI	0.094* (0.078)	-1.774*** (0.000)	0.002	0.019	2561
LVGI	0.421*** (0.003)	-2.111*** (0.000)	0.007	0.071	2561
TATA	-0.132 (0.611)	-1.678*** (0.000)	0.000	0.001	2561
m-score	0.113*** (0.002)	-1.442*** (0.000)	0.008	0.081	2561

Notes: Each variable is used as a predictor in a univariate probit where ERROR is the dependent variable. ***, **, * means significance at the 1%, 5% and 10% level. The reported p-values are from a one-sided test. The predicted sign for each coefficient is positive. Since no assumption is made regarding the sign of the constant term, the respective p-value are from a two-sided test. Prob > X^2 indicates at which level it is possible to reject the null hypothesis that the predictors have no effect on the dependent variable. The goodness of fit is expressed by the McFadden's pseudo R².

5.6.2 Univariate results: error severity and cooperation of the examined company

To test Hypothesis H2, whether error severity is associated with earnings management, I use enforcement announcements which report multiple errors with at least one affecting the core statements as a proxy for severity. Table 5.11 shows the classification of the companies with severe and non-severe errors based on the m-score. Contrary to expectations, the m-score shows that the proportion of manipulated financial statements is higher in the group with non-severe errors (37%) compared to the group with severe errors (30%). Regarding the distribution according to year and sector, there is no discernible pattern in the sample. Table 5.13 shows the variables' means and medians for the financial statements with severe and non-severe errors. Only the variable GMI is weakly significant according to the median test (p-value < 0.10), while there is no evidence that the other variables are significantly different between the two groups. Table 5.14 shows the

coefficients of the univariate probit regressions. All models have limited explanatory power (pseudo $R^2 < 0.018$), but they are all statistically significant according to the X^2 test (p-values < 0.02). Only the coefficient of AQI is weakly significant (p-value < 0.1), which might indicate that decreasing asset quality is associated with more severe errors.

Table 5.11: Classification according to the m-score per year and sector - H2

	ERRORSEV = 1				ERRORSEV = 0			
	manipulated		non-manip.		manipulated		non-manip.	
2004	0	0%	1	100%	0	0%	1	100%
2005	4	25%	12	75%	2	29%	5	71%
2006	5	56%	4	44%	4	50%	4	50%
2007	4	57%	3	43%	1	25%	3	75%
2008	3	50%	3	50%	2	25%	6	75%
2009	2	33%	4	67%	3	38%	5	63%
2010	0	0%	2	100%	1	50%	1	50%
2011	1	20%	4	80%	0	0%	1	100%
2012	0	0%	9	100%	3	100%	0	0%
2013	0	0%	3	100%	1	100%	0	0%
2014	1	25%	3	74%	0	0%	4	100%
2015	2	40%	3	60%	1	100%	0	0%
	22	30%	51	70%	18	37%	30	63%

	ERRORSEV = 1				ERRORSEV = 0			
	manipulated		non-manip.		manipulated		non-manip.	
Consumer Discretionary	2	10%	19	89%	2	18%	9	82%
Telecommunication Services	2	50%	2	50%	0	0%	1	100%
Materials	0	0%	1	100%	4	80%	1	20%
Information Technology	10	48%	11	52%	5	45%	6	55%
Consumer Staples	2	67%	1	33%	1	50%	1	50%
Health Care	2	29%	5	71%	0	0%	4	100%
Industrials	3	23%	10	77%	6	43%	8	57%
Utilities	1	33%	2	67%	0	0%	0	0%
	22	30%	51	70%	18	37%	30	63%

Notes: The variable ERRORSEV is coded 1 when the enforcement announcement reported multiple errors with at least one pertaining to the core presentation of financial statements (balance sheet and income statement), and zero otherwise. The probability of manipulation is measured through the following equation: $m_{score}_{i,t} = -4.84 + 0.920 * DSRI_{i,t} + 0.528 * GMI_{i,t} + 0.404 * AQI_{i,t} + 0.892 * SGI_{i,t} + 0.115 * DEPI_{i,t} - 0.172 * SGAI_{i,t} + 4.679 * TATA_{i,t} - 0.327 * LVGI_{i,t}$. Financial statements are classified as manipulated when the m-score is bigger than -1.78.

Table 5.12: Classification according to the m-score per year and sector - H3

	ERRORINST = 1				ERRORINST = 0			
	manipulated		non-manip.		manipulated		non-manip.	
2004	0	0%	1	100%	0	0%	1	100%
2005	0	0%	4	100%	6	32%	13	68%
2006	2	50%	2	50%	7	54%	6	46%
2007	2	100%	0	0%	3	33%	6	67%
2008	0	0%	3	100%	5	45%	6	55%
2009	2	100%	0	0%	3	25%	9	75%
2010	1	50%	1	50%	0	0%	2	100%
2011	0	0%	1	100%	1	20%	4	80%
2012	1	25%	3	75%	2	25%	6	75%
2013	1	100%	0	0%	0	0%	3	100%
2014	0	0%	2	99%	1	17%	5	83%
2015	0	0%	0	0%	3	50%	3	50%
	9	35%	17	65%	31	33%	64	67%

	ERRORINST = 1				ERRORINST = 0			
	Manipulated		Non-manipulated		Manipulated		Non-manipulated	
Consumer Discretionary	1	20%	4	79%	3	11%	24	89%
Telecommunication Services	1	33%	2	67%	1	50%	1	50%
Materials	3	100%	0	0%	1	33%	2	67%
Information Technology	1	25%	3	75%	14	50%	14	50%
Consumer Staples	1	100%	0	0%	2	50%	2	50%
Health Care	0	0%	4	100%	2	29%	5	71%
Industrials	2	33%	4	67%	7	33%	14	67%
Utilities	0	0%	0	0%	1	33%	2	67%
	9	35%	17	65%	31	33%	64	67%

Notes: The variable ERRORINST is coded 1 when the examination is conducted by the BaFin, and zero otherwise. The probability of manipulation is measured with the following equation: $mscore_{i,t} = -4.84 + 0.920 * DSRI_{i,t} + 0.528 * GMI_{i,t} + 0.404 * AQI_{i,t} + 0.892 * SGI_{i,t} + 0.115 * DEPI_{i,t} - 0.172 * SGAI_{i,t} + 4.679 * TATA_{i,t} - 0.327 * LVGI_{i,t}$. Financial statements are classified as manipulated when the m-score is bigger than -1.78.

To test hypothesis H3, whether management's lack of cooperation is associated with earnings management, I compare the examinations conducted by the FREP with those conducted by the BaFin. Table 5.12 shows that 35% of the financial statements examined by the BaFin are flagged as manipulated, which is slightly more than the 33% of those examined by the FREP. Again, the distribution of manipulators varies greatly according to year and sector, and there is no discernible pattern in the sample. Table 5.13 reports the comparison of means and medians. Contrary to expectations, financial statements examined by the FREP have a higher gross margin index (p-values < 0.01) and a higher leverage index (p-values < 0.05) than financial statements examined

by the BaFin. The depreciation index is higher for companies examined by the BaFin (p-value < 0.1). Similarly, the univariate regressions in Table 5.15 shows that the coefficient of DEPI is positive and significant at the 10% level.

Table 5.13: Descriptive statistics - H2 and H3

Variables	ERRORSEV = 1			ERRORSEV = 0			Wilcoxon test	Median test
	Mean	Median	Obs.	Mean	Median	Obs.	p-value	p-value
DSRI	1.1660	1.0528	73	1.2338	1.0430	48	0.932	0.766
GMI	0.9851	1.0059	73	0.9503	0.9909	48	0.248	0.074
AQI	1.3213	0.9821	73	1.1194	1.0073	48	0.628	0.656
SGI	1.1437	1.0870	73	1.1346	1.0709	48	0.589	0.503
DEPI	1.0003	0.9542	73	1.0791	0.9300	48	0.425	0.766
SGAI	1.0752	1.0000	73	1.2144	1.0000	48	0.438	0.909
LVGI	1.0867	1.0469	73	1.0741	1.0114	48	0.604	0.503
TATA	-0.0505	-0.0346	73	-0.0472	-0.0335	48	0.761	0.941
m-score	-1.8140	-2.1055	73	-1.9320	-2.0610	48	0.803	0.656
Variables	ERRORINST = 1			ERRORINST= 0			Wilcoxon test	Median test
	Mean	Median	Obs.	Mean	Median	Obs.	p-value	p-value
DSRI	1.1737	1.0775	26	1.1982	1.0386	95	0.840	0.351
GMI	0.8617	0.9590	26	1.0013	1.0064	95	0.002	0.000
AQI	1.1717	0.9846	26	1.2602	0.9914	95	0.833	0.693
SGI	1.1761	1.0858	26	1.1303	1.0762	95	0.900	0.962
DEPI	1.1459	1.0144	26	1.0002	0.9206	95	0.071	0.069
SGAI	0.9800	1.0000	26	1.1716	1.0000	95	0.942	0.273
LVGI	0.9868	0.9795	26	1.1076	1.0487	95	0.025	0.030
TATA	-0.0461	-0.0407	26	-0.0500	-0.0342	95	0.895	0.693
m-score	-1.9719	-2.1530	26	-1.8304	-2.0373	95	0.596	0.402

Notes: The variable ERRORSEV is coded 1 when the enforcement announcement reports multiple errors with at least one pertaining to the core presentation of financial statements (balance sheet and income statement), and zero otherwise. The variable ERRORINST is coded 1 when the examination is conducted by the BaFin, and zero otherwise. I run a Wilcoxon rank-sum test and a non-parametric equality-of-medians test. The p-values indicate at which level it is possible to reject the null hypothesis that the two samples are drawn from a population with the same median.

In conclusion, the univariate tests do not provide any evidence in support of hypothesis H2 or H3. I only find weak indications that financial statements with severe errors might be manipulated as they have a lower asset quality. Regarding financial statements of companies which refused to collaborate with the FREP, although they show a slow-down in the depreciation rate, they have a higher gross margin and a lower leverage, which is not consistent with the existence of incentives to engage in earnings management.

Table 5.14: Univariate regressions - H2

Variables	Coeff.	Constant	Prob > X²	Pseudo R²	Obs.
DSRI	-0.096 (0.709)	0.377 (0.114)	0.002	0.003	121
GMI	0.447 (0.198)	-0.171 (0.744)	0.004	0.007	121
AQI	0.198* (0.096)	0.021 (0.922)	0.011	0.018	121
SGI	0.026 (0.458)	0.232 (0.444)	0.000	0.000	121
DEPI	-0.209 (0.809)	0.479* (0.080)	0.005	0.008	121
SGAI	-0.194 (0.865)	0.482** (0.036)	0.008	0.012	121
LVGI	0.106 (0.402)	0.147 (0.757)	0.000	0.001	121
TATA	-0.133 (0.557)	0.255** (0.040)	0.000	0.000	121
m-score	0.036 (0.329)	0.329* (0.085)	0.001	0.002	121

Notes: Each variable is used as a predictor in a univariate probit where ERRORSEV is the dependent variable. ***, **, * means significance at the 1%, 5% and 10% level. The reported p-values are from a one-sided test. The predicted sign for each coefficient is positive. Since no assumption is made regarding the sign of the constant term, the respective p-values are from a two-sided test. Prob > X² indicates at which level it is possible to reject the null hypothesis that the predictors have no effect on the dependent variable. The goodness of fit is expressed by the McFadden's pseudo R².

Table 5.15: Univariate regressions - H3

Variables	Coeff.	Constant	Prob > X ²	Pseudo R ²	Obs.
DSRI	-0.034 (0.567)	-0.749*** (0.005)	0.000	0.000	121
GMI	-1.691 (0.997)	0.805 (0.172)	0.064	0.081	121
AQI	-0.084 (0.689)	-0.688*** (0.004)	0.002	0.003	121
SGI	0.112 (0.335)	-0.918*** (0.005)	0.001	0.002	121
DEPI	0.340* (0.091)	-1.150*** (0.000)	0.014	0.018	121
SGAI	-0.635 (0.883)	-0.127 (0.817)	0.023	0.029	121
LVGI	-1.232 (0.978)	0.498 (0.436)	0.038	0.048	121
TATA	0.155 (0.442)	-0.782*** (0.000)	0.000	0.000	121
m-score	-0.039 (0.670)	-0.864 (0.000)	0.002	0.002	121

Notes: Each variable is used as a predictor in a univariate probit where ERRORSEV is the dependent variable. ***, **, * means significance at the 1%, 5% and 10% level. The reported p-values are from a one-sided test. The predicted sign for each coefficient is positive. Since no assumption is made regarding the sign of the constant term, the respective p-values are from a two-sided test. Prob > X² indicates at which level it is possible to reject the null hypothesis that the predictors have no effect on the dependent variable. The goodness of fit is expressed by the McFadden's pseudo R².

5.6.3 Multivariate results

So far the univariate results have shown that error companies have higher m-scores than non-error companies and that there is a positive association between the m-score and enforcement announcements. With the help of a probit regression, I analyse if the eight ratios identified by *Beneish* (1999) are useful in predicting the existence of enforcement announcements. Prior to running the regressions, I evaluate any multicollinearity problems by estimating the Pearson's correlation coefficients. Pearson's r does not exceed 0.25 for any variable (results untabulated), indicating that there might not be multicollinearity issues.

Table 5.16 shows the results of the probit regression with and without sector and year fixed effects. Both models have significant power ($\chi^2 = 28.27$ and $\chi^2 = 81.36$ with p-values < 0.01) and descriptive validity. According to the model without fixed effects, DSRI, AQI, SGI and LVGI are positively associated with ERROR. When looking at the second model, AQI is no longer significant

and SGAI has become significant at the 10% level, providing only limited evidence that enforcement announcements are significantly associated with higher capitalisation of non-current assets and with growing costs for sales, general and administrative expenses.

These results are partially in line with the findings of the univariate analysis. They report that an increase in the days' sales in receivables compared to the previous year is associated with a higher probability of enforcement announcements, which might indicate the existence of manipulated sales and the consequent misbalance in the average collection period for receivables. Further, the positive association between the leverage index and the existence of enforcement announcements might be consistent with financially distressed companies being more likely to resort to earnings management. There is only limited indication that the probability of having enforcement announcements increases for companies with a disproportionate increase in sales, general and administrative expenses and with a higher capitalisation of non-current assets other than PP&E. Contrary to findings of the univariate tests, I observe here that quickly growing companies are more likely to have enforcement announcements, which might indicate that earnings management take place through manipulated sales.

Table 5.16: Multivariate regressions with and without year and sector fixed effects - HI

ERROR	Constant	DSRI	GMI	AQI	SGI	DEPI	SGAI	LVGI	TATA	Pseudo R²	X²
Coeff.	-2.2508***	0.2634***	-0.4378	0.1230**	0.2373**	-0.0863	0.0693	0.3036**	-0.3309	0.029	28.27
(p-value)	(0.000)	(0.001)	(0.975)	(0.025)	(0.026)	(0.756)	(0.167)	(0.026)	(0.751)		(0.000)
ERROR with year and sector fixed effects											
Coeff.	-2.2484***	0.2390***	-0.5648	0.0817	0.2895**	-0.0690	0.1066*	0.3288**	-0.3907	0.083	81.36
(p-value)	(0.000)	(0.004)	(0.991)	(0.104)	(0.012)	(0.706)	(0.081)	(0.021)	(0.777)		(0.000)

Notes: The number of observations is 2561. ***, **, * means significance at the 1%, 5% and 10% level. The reported p-values are from a one-sided test. The predicted sign for each variable is positive. Since no assumption is made regarding the sign of the constant term, the reported p-value is from a two-sided test. Prob > X² indicates at which level it is possible to reject the null hypothesis that the predictors have no effect on the dependent variable. The goodness of fit is expressed by the McFadden's pseudo R².

To provide further insights into the hypotheses H2 and H3, I run two separated regressions based on the data for the error group. The sample includes 121 firm-year observations, which means the results might be affected by the small size of the sample. In fact, due to the limited number of observations, some sector and year dummies suffer from multicollinearity. Therefore, I run only a probit regression without year and sector fixed effects.

Consistently, with the univariate test, which shows that only AQI is (weakly) associated with ERRORSEV, I find here that the coefficient of AQI is positive and weakly significant, indicating that decreasing asset quality is associated with a higher probability of the announcement reporting a severe error. However, the model is not a good fit ($\chi^2 = 3.05$ and p-value = 0.763) and all other coefficients are not significant, which leads to the rejection of hypothesis H2. This finding is partially in line with Böcking/Gros/Worret (2015), who find no significant difference in terms of earnings management between error companies with core statement errors and companies with disclosure errors¹⁶⁹.

In the second regression, even though the model has descriptive validity ($\chi^2 = 20.97$ and p-value = 0.007), none of the coefficients is significant, not even DEPI which according to the univariate probit was positively associated with ERRORINST. Therefore, hypothesis H3 must be rejected. This finding contrasts with that of Pasch (2017), who determines that the cooperation of the company's management, in terms of the quality of information they provide and their timely response, is negatively related to the existence of enforcement findings¹⁷⁰. However, unlike this study, Pasch (2017) measures the degree of cooperativity based on non-public data of the FREP and uses data, which are not publicly available, regarding examinations which ended without enforcement findings. It is possible that assuming the lack of cooperation when the examination is conducted by the BaFin is not a good proxy for the level of cooperation. Also, as the sample to test hypotheses H2 and H3 only includes 121 firm-year observations, the results might be affected by small sample issues.

Table 5.17: Multivariate regressions without year and sector fixed effects - H2 and H3

ERRORSEV	Constant	DSRI	GMI	AQI	SGI	DEPI	SGAI	LVGI	TATA	Pseudo R2	chi2
Coeff.	-0.0288	0.0214	0.3114	0.2206*	-0.0789	-0.2333	-0.2371	0.2640	-0.2234	0.0305	4.953
(p-value)	(0.977)	(0.465)	(0.293)	(0.081)	(0.602)	(0.821)	(0.872)	(0.282)	(0.585)		(0.763)
ERRORINST											
Coeff.	3.6177**	0.0950	-2.2819	-0.1130	0.1192	0.2216	-0.6828	-1.8287	-0.3337	0.1666	20.974
(p-value)	(0.017)	(0.379)	(0.999)	(0.725)	(0.373)	(0.227)	(0.892)	(0.995)	(0.596)		(0.007)

Notes: The number of observations is 121. ***, **, * means significance at the 1%, 5% and 10% level. The reported p-values are from a one-sided test. The predicted sign for each variable is positive. Since no assumption is made regarding the sign of the constant term, the reported p-value is from a two-sided test. Prob > X² indicates at which level it is possible to reject the null hypothesis that the predictors have no effect on the dependent variable. The goodness of fit is expressed by the McFadden's pseudo R².

¹⁶⁹ Böcking/Gros/Worret (2015), p. 481.

¹⁷⁰ Pasch (2017), pp. 21, 28.

5.7 Conclusion

The Financial Reporting Compliance Act established in 2004 the German enforcement system. The capacity of the enforcement system to enhance the quality of financial statements depends also on its ability to identify companies engaging in earnings management¹⁷¹. However, enforcement announcements do not provide information on the determinants of errors, failing to inform about the extent of earnings management among examined companies and about the efficiency of the German enforcement system in detecting manipulated financial statements.

In this paper I have analysed the relation between erroneous financial statements, identified with enforcement announcements, and earnings management. Consistent with the existence of earnings management, the univariate results show that error companies have higher m-scores than non-error companies and that there is a positive association between the m-score and enforcement announcements. These findings are in line with prior studies, which find evidence that the German enforcement system is efficient in findings manipulators¹⁷².

By employing several financial statement ratios, I am able to shed some light on the determinants of erroneous financial reporting. The results of the multivariate probit regression seems to indicate that the model of *Beneish* (1999) is, at least partially, useful in explaining the existence of enforcement announcements among companies falling under the enforcement regime and that an abnormal increase in the days' sales in receivables, a raise in financial pressure and an unusual growth in sales might be useful in predicting the existence of enforcement announcements. These findings are consistent with financial pressure being a motive for and with manipulated sales being an indication of earnings management. Conversely, the analysis provides only weak evidence that earnings management is conducted with the improper capitalisation of non-current assets other than PP&E and that the probability of having enforcement announcements is positively associated with a disproportionate increase in sales, general and administrative expenses.

Finally, I analysed the information content of enforcement announcements, in terms of the severity of error and the lack of cooperation of the company's managers, to determine if they can be interpreted as signals for ongoing earnings management. The results are inconclusive, not

¹⁷¹ Böcking/Gros/Worret (2015), p. 433; Strohmenger (2014), pp. 273-274.

¹⁷² Böcking/Gros/Worret (2015), p. 482, Strohmenger (2013), p. 28, Strohmenger (2014), p. 296, and Hoehn/Strohmenger (2013), p. 3.

providing any evidence that financial statements with more severe errors or where managers are uncooperative are more likely to have manipulated earnings.

This study suffers from a significant limitation regarding the ability of the m-score to correctly predict earnings management. I employed in this study the original model of *Beneish* (1999). However, as the model was estimated on a sample of companies that manipulated their earnings between 1982 and 1992, the question arises whether it should be re-estimated using current data and updated considering the most recent literature on fraud and country-specific aspects. Future studies willing to update the model should make sure that companies which have definitely manipulated their earnings are recognised and included in the estimation sample.

6. Summary and conclusion

In this dissertation, I carried out four studies, each focusing on issues related to the oversight of financial information in Germany. Some major accounting scandals as well as the financial crisis have shown the need for further improving the integrity and reliability of financial reporting, which is why some regulative actions have been undertaken, both at European and at national level. To enhance the quality of financial statements, the EU has promoted the institution of national enforcement systems, which oversee the compliance of financial information with the applicable reporting framework, and has issued, in April 2014, Directive 2014/56/EU, which amends the existing Directive 2006/43/EC, and Regulation (EU) No 537/2014, which applies to the statutory audit of PIEs.

In chapter 2, I examined, together with Renate Hecker, whether certain limitations introduced by the EU-regulation are potentially able to improve audit quality. In chapter 3, I built on and complemented the descriptive and univariate analyses carried out in chapter 2, by running multivariate analyses and by performing several sensitivity tests to verify the robustness of the findings. Regarding the provision of NAS, the results of the two studies do not show that companies buying NAS are more likely to have enforcement announcements. Although there is no evidence that other assurance services are associated with audit quality, there are indications that the provision of other non-audit services and of tax services is associated with higher audit quality. It follows that both the descriptive and the multivariate analyses do not support the necessity of further restricting the provision of NAS in Germany. On the contrary, they suggest that tax services, and possibly also other non-audit services, might have the potential of increasing audit quality by generating additional client-specific knowledge. For this reason, the decision of the German regulator not to extend the black list and to allow the provision of certain tax services seems appropriate.

Regarding the introduction of the cap, the analyses do not show that companies exceeding the 70% threshold have lower audit quality, thus do not support the argument that the introduction of the cap might improve auditor independence. Also, there are uncertainties on whether the cap will have a significant impact among PIEs, as in the examined period, prior to the introduction of the black list, most companies were already paying fees below the cap or did not qualify for its application. Given the potentially limited application as well as the lack of evidence supporting the choice of

the 70% threshold, it remains unclear if and to what extent audit quality might benefit from the introduction of the cap.

Similar concerns exist in relation to limiting the proportion of revenue an auditor can obtain from a client. The analysis shows that, partially due to the existing regulation, it is rather uncommon that among publicly traded companies an auditor obtains more than 15% of his revenue from a client, which indicates a potentially limited impact of this regulatory measure. While the descriptive analysis shows that companies with a fee-to-revenue ratio higher than 15% are more likely to have enforcement announcements, the multivariate analysis does not provide any evidence in this regard. The partial indications obtained that exceeding the 15% threshold might be negatively associated with audit quality, together with the theoretical considerations that client importance might impair auditor independence, seem to suggest that it might be wise to limit the financial dependence of an auditor on a client, although setting the limit at 15% might be too high.

Regarding the length of the audit engagement, the descriptive analysis carried out in chapter 2 does not show that audit quality is lower in the engagement years which follow an auditor change. This finding does not provide evidence against limiting the length of the audit engagement. However, it also does not provide support for mandatory auditor rotation, as the analysis does not investigate whether audit quality is negatively associated with long engagements. It follows that it might be possible that the auditor develops a close relationship with the client after working together for many years, which might affect his objectivity and professional scepticism.

In chapter 4, I tried to provide further answers to the concerns expressed by the European regulator regarding the provision of certain NAS and the level and structure of the fees paid by audited companies to their statutory auditor. Building on the theoretical auditor-client dynamic outlined by *DeAngelo* (1981a), I examined the relation between audit quality and the provision of NAS in a multi-period framework¹. I observe that enforcement announcements are positively associated with future fees and negatively associated with current NAS fees, which is consistent with the presence of an economic bond due to the expected quasi-rents and with the existence of knowledge spillovers. The results indicate that the provision of NAS might generate additional knowledge, which might increase the ability of the auditor to detect errors. Considering this evidence, there seems to be no need for further limiting the provision of NAS, as this might reduce audit quality.

¹ DeAngelo (1981a), p. 115.

Also, as there is only limited evidence that the expected provision of future NAS is associated with enforcement findings, it is unclear whether further restricting the provision of NAS or putting a cap on NAS fees might help to increase auditor independence.

In chapter 5, I analysed the relation between erroneous financial statements, identified with enforcement announcements, and earnings management, using the m-score. In line with prior studies, the results suggest that the German enforcement system might be efficient in finding companies engaging in earnings management². By employing several financial statement ratios, I have been able to shed some light on the determinants of erroneous financial reporting. The findings obtained are consistent with financial pressure being a motive for earnings management and with manipulated sales being an indication of earnings management. Conversely, there is only weak evidence that earnings management is conducted with the improper capitalisation of non-current assets other than PP&E and that the probability of having enforcement announcements is positively associated with a disproportionate increase in sales, general and administrative expenses. The analysis of the information content of enforcement announcements, in terms of the severity of error and the lack of cooperation of the company's managers, is inconclusive, not providing any indications that they can be interpreted as signals for ongoing earnings management.

Concluding, the analysis of certain issues regarding the oversight of financial information in Germany shows that the two-tier enforcement system is potentially able to find and sanction companies engaging in earnings management, which means that its establishment might contribute to increasing the quality of financial information. Conversely, there is only partial evidence that restricting the provision of certain NAS, setting a cap on NAS fees as well as on the proportion of fees an auditor can obtain from a client, and limiting the length of audit engagements are effective in strengthening the role of the statutory auditor and in enhancing the quality of financial statements. It should be noted that, as these studies have been conducted in Germany, the findings might not be generalised or apply to other countries due to different enforcement systems and institutional characteristics, such as auditor's liability, accounting regime, ownership structure and NAS-related characteristics³.

² Böcking/Gros/Worret (2015), p. 482, Strohmenger (2013), p. 28, Strohmenger (2014), p. 296, and Hoehn/Strohmenger (2013), p. 3.

³ For other limitations, see section 2.11, section 3.6, section 4.5 and section 5.7.

References

- Anh, N. H. and N. H. Linh. 2016. Using the M-score Model in detecting earnings management: Evidence from non-financial Vietnamese listed companies. *VNU Journal of Science: Economics and Business* 32(2): 14-23.
- Allison, P. D. 2012. *Logistic regression using SAS: theory and application*. 2nd edition. Cary, NC: SAS Institute Inc.
- Antle, R., E. Gordon, G. Narayanamoorthy, and L. Zhou. 2006. The joint determination of audit fees, non-audit fees, and abnormal accruals. *Review of Quantitative Finance and Accounting* 27(3): 235-266.
- Armstrong, C. S., A. D. Jagolinzer, and D. F. Larcker. 2010. Chief executive officer equity incentives and accounting irregularities. *Journal of Accounting Research* 48(2): 225-271.
- Arruñada, B. 1999. *The economics of audit quality. Private incentives and the regulation of audit and non-audit services*. Boston: Kluwer Academic Publishers.
- BaFin. 2013. Jahresbericht der BaFin. Available at:
https://www.bafin.de/DE/PublikationenDaten/Jahresbericht/jahresbericht_node.html
- BaFin. 2014. Jahresbericht der BaFin. Available at:
https://www.bafin.de/DE/PublikationenDaten/Jahresbericht/jahresbericht_node.html
- BaFin. 2015. Jahresbericht der BaFin. Available at:
https://www.bafin.de/DE/PublikationenDaten/Jahresbericht/jahresbericht_node.html
- BaFin. 2016. Jahresbericht der BaFin. Available at:
https://www.bafin.de/DE/PublikationenDaten/Jahresbericht/jahresbericht_node.html
- BaFin. 2017. Jahresbericht der BaFin. Available at:
https://www.bafin.de/DE/PublikationenDaten/Jahresbericht/jahresbericht_node.html
- BaFin. 2018. Emittentenleitfaden der BaFin. Available at:
https://www.bafin.de/SharedDocs/Downloads/DE/Leitfaden/WA/dl_emittentenleitfaden_einleitung.html?nn=11407966
- Beck, P. J., T. J. Frecka, and I. Solomon. 1988. A model of the market for MAS and audit services: Knowledge spillovers and auditor-auditee bonding. *Journal of Accounting Literature* 7(1): 50-64.
- Beneish, M. D. 1997. Detecting GAAP violation: implications for assessing earnings management among firms with extreme financial performance. *Journal of Accounting and Public Policy* 16(3): 271-309.
- Beneish, M. D. 1999. The detection of earnings manipulation. *Financial Analysts Journal* 55(5): 24-36.

- Beneish, M. D. 2001. Earnings management: a perspective. *Managerial Finance* 27(12): 3-17.
- Beneish, M. D., C. M. C., Lee, and D. C. Nichols. 2013a. Earnings manipulation and expected returns. *Financial Analysts Journal* 69(2): 57-82.
- Beneish, M. D., C. M. C., Lee, and D. C. Nichols. 2013b. Earnings manipulation and expected returns: Author response. *Financial Analysts Journal* 69(5): 14.
- Beneish, M. D. and D. C. Nichols. 2007. The predictable cost of earnings manipulation. Working paper. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1006840.
- Beneish, M. D., C. Nichols, and C. M. C., Lee. 2011. To catch a thief: Can forensic accounting help predict stock returns? Working paper. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1903593.
- Berger, A. 2010. The development and status of enforcement in the European Union. *Accounting in Europe* 7(1): 15-35.
- Beyhs, O., E. Kühne, and H. Zülch. 2012. Abschlussprüfung und DPR-Verfahren - Darstellung und Würdigung der Verfahrensunterschiede. *Die Wirtschaftsprüfung* 65(12): 650-660.
- Bischof, S. 2006. Anhangangaben zu den Honoraren für Leistungen des Abschlussprüfers. *Die Wirtschaftsprüfung* 59(11): 705-713.
- Blay, A. D. and M. A. Geiger. 2013. Auditor fees and auditor independence: Evidence from going concern reporting decisions. *Contemporary Accounting Research* 30(2): 579-606.
- Blaylock, B., F. B. Gaertner, and T. Shevlin. 2017. Book-tax conformity and capital structure. *Review of Accounting Studies* 22(2): 903-932.
- Bockmann, R. 2012. *Internationale Koordinierung nationaler Enforcement-Aktivitäten. Eine kritische Analyse unter besonderer Berücksichtigung der Deutschen Prüfstelle für Rechnungslegung*. Wiesbaden: Springer Gabler.
- Böcking, H. J., M. Gros, and D. Worret. 2015. Enforcement of accounting standards: how effective is the German two-tier system in detecting earnings management? *Review of Managerial Science* 9(3): 431-485.
- Bose, R. and T. Lilienbecker. 2019. Praktische Auslegungsfragen zur Begrenzung von Nichtprüfungsleistungen. *Betriebs-Berater* 74(13): 746-749.
- Brambor, T., W. R. Clark, and M. Golder. 2006. Understanding interaction models: Improving empirical analyses. *Political Analysis* 14(1): 63-82.
- Buis, M. L. 2010. Stata tip 87: Interpretation of interactions in nonlinear models. *The Stata Journal* 10(2): 305-308.
- Cahan, S., D. Emanuel, D. Hay, and N. Wong. 2008. Non-audit fees, long-term auditor-client relationships and earnings management. *Accounting and Finance* 48(2): 181-207.

Caliendo, M. and S. Kopeinig. Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys* 22(1): 31-72.

Cassell, C. A., J. N. Myers, L. A. Myers, and T. A. Seidel. 2016. Does auditor tenure impact the effectiveness of auditors' response to fraud risk? Working paper. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2448680.

Castillo-Merino D., J. Garcia-Blandon, and M. Martinez-Blasco. 2018. Auditor independence, current and future NAS fees and audit quality: Were European regulators right? Working paper (accepted for publication in the *European Accounting Review*). Available at: https://www.researchgate.net/profile/David_Castillo-Merino/publication/331217835_Auditor_Independence_Current_and_Future_NAS_Fees_and_Audit_Quality_Were_European_Regulators_Right/links/5c6d5ec84585156b570be5eb/Auditor-Independence-Current-and-Future-NAS-Fees-and-Audit-Quality-Were-European-Regulators-Right.pdf.

Causholli, M., D. J. Chambers, and J. L. Payne. 2014. Future nonaudit service fees and audit quality. *Contemporary Accounting Research* 31(3): 681-712.

Cheng, S., L. Martin, and R. E. Werum. 2007a. Adult social capital and track placement of ethnic groups in Germany. *American Journal of Education* 114(1): 41-74.

Cheng, S., L. Martin, and R. E. Werum. 2007b. Appendix accompanying "Adult social capital and track placement of ethnic groups in Germany". *American Journal of Education* 114(1): 41-74. Available at: <https://web.uconn.edu/simoncheng/AJE114-1.pdf>

Chung H. and S. Kallapur. 2003. Client importance, Nonaudit services, and abnormal accruals. *The Accounting Review* 78(4): 931-955.

Cleff, T. 2019. *Applied Statistics and Multivariate Data Analysis for Business and Economics*. Cham: Springer.

Corder, G. W., D. I. Foreman. 2009. *Nonparametric statistics for non-statisticians*. A step-by-step approach. New Jersey: John Wiley & Sons.

Corsi, C., D. Di Bernardino, and T. Di Cimbrini. 2015. Beneish M-score and detection of earnings management in Italian SMEs. *Ratio Mathematica* 26(28): 65-83.

Cram, D. P., V. Karan, and I. Stuart. 2009. Three threats to validity of choice-based and matched-sample studies in accounting research. *Contemporary Accounting Research* 26(2): 477-516.

DeAngelo, L. E. 1981a. Auditor independence, 'low balling', and disclosure regulation. *Journal of Accounting and Economics* 3(2): 113-127.

DeAngelo, L. E. 1981b. Auditor size and audit quality. *Journal of Accounting and Economics* 3(3): 183-199.

- Dechow, P. M. and D. J. Skinner. 2000. Earnings management: reconciling the views of accounting academics, practitioners, and regulators. *Accounting Horizons* 14(2): 235-250.
- Dechow, P. M., W. Ge, C. R. Larson, and R. G. Sloan. 2011. Predicting material accounting misstatements. *Contemporary Accounting Research* 28(1): 17-82.
- Dechow, P.M., W. Ge, and C. Schrand. 2010. Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting and Economics* 50(2-3): 344-401.
- Dechow, P. M., R. G. Sloan, and A. P. Sweeney. 1996. Causes and consequences of earnings manipulation: An analysis of firms subject to enforcement actions by the SEC. *Contemporary Accounting Research* 13(1): 1-36.
- DeFond, M. L. and J. Jiambalvo. 1991. Incidence and circumstances of accounting errors. *The Accounting Review* 66(3): 643-655.
- DeFond, M. L., C. Y. Lim, and Y. Zang. 2016. Client conservatism and auditor-client contracting. *The Accounting Review* 91(1): 69-98.
- DeFond, M. L., K. Raghunandan, and K.R. Subramanyam. 2002. Do non-audit service fees impair auditor independence? Evidence from going concern audit opinions. *Journal of Accounting Research* 40(4): 1247-1274.
- DeFond, M. and J. Zhang. 2014. A review of archival auditing research. *Journal of Accounting and Economics* 58(2-3): 275-326.
- Diemers, M. and N. B. Homfeldt. 2016. Änderungen des WpHG infolge der EU-Transparenzrichtlinie. Eine überblicksartige Darstellung und ausgewählte Problemfelder. *Praxis der internationalen Rechnungslegung* (2): 45-52.
- Dobler, M. 2014. Auditor-provided non-audit services in listed and private family firms. *Managerial Auditing Journal* 29(5): 427-454.
- DPR. 2005. Grundsätze für die stichprobenartige Prüfung gemäß § 342b Abs. 2 Satz 3 Nr. 3 HGB vom 05.09.2005. Available at: https://www.frep.info/docs/rechtliche_grundlagen/20050905_grundsaeetze_stichprobenartige_pruefung.pdf.
- DPR. 2006. Tätigkeitsbericht für den Zeitraum vom 01.07 bis 31.12.2005. Available at: https://www.frep.info/docs/jahresberichte/2005/2005_tb.pdf.
- DPR. 2009. Grundsätze für die stichprobenartige Prüfung gemäß § 342b Abs. 2 Satz 3 Nr. 3 HGB vom 20.04.2009. Available at: https://www.frep.info/docs/rechtliche_grundlagen/20090420_grundsaeetze_stichprobenartige_pruefung.pdf.

- DPR. 2015. 10 Jahre Bilanzkontrolle in Deutschland (2005 bis 2015). Available at: https://www.frep.info/docs/dpr_10_jahre/dpr_jubilaebumsbroschuere.pdf
- DPR. 2016. Grundsätze für die stichprobenartige Prüfung gemäß § 342b Abs. 2 Satz 3 Nr. 3 HGB vom 15.12.2016. Available at: https://www.frep.info/docs/rechtliche_grundlagen/20161215_grundsaeetze_stichprobenartige_pruefung.pdf.
- Ebke, W. F. 2013. HGB § 319a Besondere Ausschlussgründe bei Unternehmen von öffentlichem Interesse, in Münchener Kommentar, 3rd edition, Munich.
- Ebner, G., J. Hottmann, and H. Zülch. 2017. Error announcement, auditor turnover, and earnings management – Evidence from Germany. *Corporate Ownership & Control* 14(3): 122-151.
- EC. 2014. Q&A – Implementation of the new statutory audit framework. Available at: <https://www.idw.de/blob/84924/86111f8267274d7cbd4f665444494efd/down-eu-kommission-fragen-u-antworten-sept-2014-data.pdf>
- EC. 2017. Feedback statement on the public consultation on the operations of the European Supervisory Authorities having taken place from 21 March to 16 May 2017. Available at: https://ec.europa.eu/info/sites/info/files/2017-esas-operations-summary-of-responses_en.pdf.
- Eilifsen, A., R. Quick, F. Schmidt, and S. Umlauf. 2018. Investors' perceptions of nonaudit services and their type in Germany: The financial crisis as a turning point. *International Journal of Auditing* 22(2): 298-316.
- Eisenhardt, K. M. 1989. Agency theory: An assessment and review. *Academy of Management Review* 14(1): 57-74.
- Ellrot, H. 2006. HGB § 285 Sonstige Pflichtangaben, in Beck'scher Bilanz-Kommentar, 6th edition. Munich.
- Ellrot, H. 2006. HGB § 314 Sonstige Pflichtangaben, in Beck'scher Bilanz-Kommentar, 6th edition. Munich.
- Ernst, C., T. Amann, and P. Kraus. 2017. Deutsche Prüfstelle für Rechnungslegung eV (DPR). Vorgehensweise und Prüfungsschwerpunkte. In W. Funk and J. Rossmann (eds.). *Internationale Rechnungslegung und Internationales Controlling: Herausforderungen – Handlungsfelder - Erfolgspotenziale* (pp. 235-248). Wiesbaden: Springer Fachmedien.
- Ernstberger, J., J.-M. Hitz, and M. Stich. 2012. Why do firms produce erroneous IFRS financial statements? Working paper. Available at SSRN: <https://ssrn.com/abstract=2060328>
- Ernstberger, J., M. Stich, and O. Vogler. 2012. Economic consequences of accounting enforcement reforms: the case of Germany. *European Accounting Review* 21(2): 217-251.

ESMA. 2017. Peer review on guidelines on enforcement of financial information. Peer Review Report. Available at: https://www.esma.europa.eu/sites/default/files/library/esma42-111-4138_peer_review_report.pdf.

Ettredge, M., S. Scholz, K. R. Smith, and L. Sun. 2010. How do restatements begin? Evidence of earnings management preceding restated financial reports. *Journal of Business Finance & Accounting* 37(3-4): 332-355.

Ewert, R. 2004. Audit regulation, audit quality, and audit research in the post-Enron era: An analysis of nonaudit services. In Leuz C., D. Praff, and A. Hopwood (eds). *The economics and politics of accounting: International perspectives on research trends, policy, and practice* (pp. 239-263). New York: Oxford University Press.

Feruleva, N. V. and M. A. Shtefan. 2017. Detecting financial statements fraud: The evidence from Russia. *Journal of Corporate Finance Research* 12(2): 32-45.

Fiallo, C. and R. Hecker. 2019a. Ausgewählte Neuregelungen der EU-Verordnung zur Abschlussprüfung und Prüfungsqualität (Teil 1) - Eine empirische Bestandsaufnahme bei deutschen Unternehmen mit DPR-Fehlerfeststellungen. *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung (KoR)* 19(5): 225-234.

Fiallo, C. and R. Hecker. 2019b. Ausgewählte Neuregelungen der EU-Verordnung zur Abschlussprüfung und Prüfungsqualität (Teil 2) - Eine empirische Bestandsaufnahme bei deutschen Unternehmen mit DPR-Fehlerfeststellungen. *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung (KoR)* 19(6): 290-297.

Fiallo, C. and R. Hecker. 2019c. Ausgewählte Neuregelungen der EU-Verordnung zur Abschlussprüfung und Prüfungsqualität (Teil 3) - Eine empirische Bestandsaufnahme bei deutschen Unternehmen mit DPR-Fehlerfeststellungen. *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung (KoR)* 19(7-8): 343-348.

Franceschetti, B. M. and C. Koschtial. 2013. Do bankrupt companies manipulate earnings more than the non-bankrupt ones? *Journal of Finance and Accountancy* 12: 1-23.

Francis, J. R. 2004. What do we know about audit quality? *The British Accounting Review* 36(4): 345-368.

Francis, J. R. 2011. A framework for understanding and researching audit quality. *Auditing: A Journal of Practice & Theory* 30(2): 125-152.

Frank, K. A. 2000. Impact of a confounding variable on a regression coefficient. *Sociological Methods & Research* 29(2): 147-194.

FREP. 2006. Activity Report for the period from 1 July to 31 December 2005. Available at: https://www.frep.info/docs/jahresberichte/2005/2005_tb_en.pdf.

FREP. 2007. Annual activity Report for the period from January 1 through December 31, 2006. Available at: https://www.frep.info/docs/jahresberichte/2006/2006_tb_en.pdf.

FREP. 2008. Annual Activity Report 2007. Available at: https://www.frep.info/docs/jahresberichte/2007/2007_tb_en.pdf.

FREP. 2009. Annual Activity Report 2008. Available at: https://www.frep.info/docs/jahresberichte/2008/2008_tb_en.pdf.

FREP. 2010. Annual Activity Report 2009. Available at: https://www.frep.info/docs/jahresberichte/2009/2009_tb_en.pdf.

FREP. 2011. Annual Activity Report 2010. Available at: https://www.frep.info/docs/jahresberichte/2010/2010_tb_en.pdf.

FREP. 2012. Annual Activity Report 2011. Available at: https://www.frep.info/docs/jahresberichte/2011/2011_tb_en.pdf.

FREP. 2013. Annual Activity Report 2012. Available at: https://www.frep.info/docs/jahresberichte/2012/2012_tb_en.pdf.

FREP. 2014. Annual Activity Report 2013. Available at: https://www.frep.info/docs/jahresberichte/2013/2013_tb_en.pdf.

FREP. 2015a. Annual Activity Report 2014. Available at: https://www.frep.info/docs/jahresberichte/2014/2014_tb_en.pdf.

FREP. 2015b. Case-related ex-ante query. Available at: https://www.frep.info/docs/fallbezogene_voranfragen/info_fallbezogene_voranfragen_en.pdf.

FREP. 2016. Annual Activity Report 2015. Available at: https://www.frep.info/docs/jahresberichte/2015/2015_tb_en.pdf.

FREP. 2017. Annual Activity Report 2016. Available at: http://www.frep.info/docs/jahresberichte/2016/2016_tb_en.pdf.

FREP. 2018a. Activity report 2017. Available at: https://www.frep.info/docs/jahresberichte/2017/2017_tb_en.pdf.

FREP. 2018b. Information on the examination process of the Financial Reporting Enforcement Panel (FREP). Available at: https://www.frep.info/docs/pruefverfahren/info_ablauf_pruefverfahren_en.pdf.

Frey L., M. Möller, and R. Weinzierl. 2016. § 37q - WpHG – Fehlerveröffentlichungen und Wechsel der Abschlussprüfungsgesellschaft. Eine empirische Untersuchung. *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung (KoR)* 16(12): 563-574.

Fuhrmann, E.-C. 2016. *(Future) non-audit fees and audit quality – An empirical investigation for Germany*. Master's thesis. University of Tübingen.

- Gahlen, D. and A. Schäfer. 2006. Bekanntmachung von fehlerhaften Rechnungslegungen im Rahmen des Enforcementverfahrens: Ritterschlag oder Pranger? *Betriebs-Berater* 61(30): 1619–1623.
- Garcia-Blandon, J., J. M. Argilés-Bosch, and D. Ravenda. 2019. Audit firm tenure and audit quality: A cross-European study. *Journal of International Financial Management & Accounting* Version of Record online: 26.04.2019: 1-30.
- Gros, M. 2016. Non-audit services, audit quality and enforcement – evidence from German enforcement examinations. *International Journal of Economics and Accounting* 7(3): 218-249.
- Gros, M. and D. Worret. 2014. The challenge of measuring audit quality: some evidence. *International Journal of Critical Accounting* 6(4): 345-374.
- Grottel, B. 2016. HGB § 314 Sonstige Pflichtangaben, in Beck'scher Bilanz-Kommentar, 10th edition, Munich.
- Grottel, B. 2016. HGB § 342b Prüfstelle für Rechnungslegung, in Beck'scher Bilanz-Kommentar, 10th edition, Munich.
- Grottel, B. 2018. HGB § 285 Sonstige Pflichtangaben, in Beck'scher Bilanz-Kommentar, 11th edition, Munich.
- Grottel, B. 2018. HGB § 314 Sonstige Pflichtangaben, in Beck'scher Bilanz-Kommentar, 11th edition, Munich.
- Grove, H. and E. Basilico. 2008. Fraudulent financial reporting detection: Key ratios plus corporate governance factors. *International Studies of Management & Organization* 38(3): 10-42.
- Grove, H. and T. Cook. 2004. Lessons for auditors: Quantitative and qualitative red flags. *Journal of Forensic Accounting* 5(1): 131-146.
- Häfele, S., and M. Riediger. 2015. Consequences of error announcements on industry peers: capital market reaction and financial reporting decisions. Working paper. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2541903.
- Healy, P. M. and J. M. Wahlen. 1999. A review of the earnings management literature and its implications for standard setting. *Accounting Horizons* 13(4): 365-383.
- Hecker, R. and A. Wild. 2012. The market effects of the German two-tier enforcement of financial reporting. Tübinger Diskussionsbeitrag Nr. 334. Available at: <https://publikationen.uni-tuebingen.de/xmlui/handle/10900/47996>.
- Hennes, K. M., A. J. Leone, and B. P. Miller. 2008. The importance of distinguishing errors from irregularities in restatement research: the case of restatements and CEO/CFO turnover. *The Accounting Review* 83(6): 1487-1519.

Herzig, N., and C. Watrin. 1995. Obligatorische Rotation des Wirtschaftsprüfers – ein Weg zur Verbesserung der externen Unternehmenskontrolle? *Zeitschrift für betriebswirtschaftliche Forschung* 47(9): 725-804.

Hitz, J.-M., J. Ernstberger, and M. Stich. 2012. Enforcement of accounting standards in Europe: Capital-market-based evidence for the two-tier mechanism in Germany. *European Accounting Review* 21(2): 253-281.

Hitz, J.-M., P. Löw, and M. Solka. 2013. Determinants of audit delay in a mandatory IFRS setting. *Die Betriebswirtschaft* 73(4): 293-305.

Hoehn, B. and M. Strohmenger. 2013. The effects of earnings management on enforcement releases and their recognition in audit fees. Working paper. Available at SSRN: <https://ssrn.com/abstract=2302680>.

Hohenfels, D. and R. Quick. 2018. Non-audit services and audit quality: evidence from Germany. *Review of Managerial Science* OnlineFirst 25 October 2018. Available at: <https://link.springer.com/article/10.1007/s11846-018-0306-z>.

Holm, C. 2016. The new statutory audit framework in Europe: Implications for the provisions of non-audit services. Working paper. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2823477.

Höltken, M., S. Jana, and H. Zülch. 2016. The association between executive turnover and financial misreporting: Evidence from Germany. Working paper. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2578377.

Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. *Applied logistic regression*. 3rd edition. New Jersey: John Wiley & Sons, Inc.

Hubbard, T. D., D. M. Christensen, and S. D. Graffin. 2017. Higher highs and lower lows: The role of corporate social responsibility in CEO dismissal. *Strategic Management Journal* 38(11): 2255-2265.

IDW. 2018a. *Positionspapier zu Inhalten und Zweifelsfragen der EU-Verordnung und der Abschlussprüferrichtlinie*. 4th edition (23/05/2018). Available at: <https://www.idw.de/blob/86498/ea346a862c94cf8619af57187ea153ee/down-positionspapier-zweifelsfragen-data.pdf>

IDW. 2018b. *Positionspapier zu Nichtprüfungsleistungen des Abschlussprüfers*. 4th edition (05/11/2018). Available at: <https://www.idw.de/idw/idw-aktuell/idw-positionspapier-zu-nichtpruefungsleistungen-des-abschlusspruefers/98216>.

IDW. 2019. *WP-Handbuch. Wirtschaftsprüfung & Rechnungslegung*. 16th edition. Düsseldorf: IDW Verlag GmbH.

IESBA. 2018. *Handbook of the International Code of Ethics for Professional Accountants*. New York: International Federation of Accountants.

Jackson, A. B. 2018. Discretionary accruals: Earnings management... or not? *Abacus* 54(2): 136-153.

Jones, K. L., G. V. Krishnan, and K. D. Melendrez. 2008. Do models of discretionary accruals detect actual cases of fraudulent and restated earnings? An empirical analysis. *Contemporary Accounting Research* 25(2): 499-531.

Kamal, M. E. M., M. F. M. Salleh, and A. Ahmad. 2016. Detecting financial statement fraud by Malaysian public listed companies: The reliability of the Beneish M-Score model. *Jurnal Pengurusan (UKM Journal of Management)* 35(46): 23-32.

Klås, E. and J. R. Werner. 2014. Financial statement error findings and auditor reputation: Evidence from the German enforcement regime. Working paper. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2417878

Köhler, A. G., K. U. Marten, R. Meuwissen, and R. Quick. 2013. Editorial - Research on audit quality. *Die Betriebswirtschaft* 73(4): 241-253.

Köhler, A. G., K. U. Marten, N. V. Ratzinger-Sakel, and M. Wagner. 2010. Prüfungshonorare in Deutschland – Determinanten und Implikationen. *Zeitschrift für Betriebswirtschaft* 80(1): 5-29.

Köhler, A. G. and N. V. Ratzinger-Sakel. 2012. Audit and non-audit fees in Germany - The impact of audit market characteristics. *Schmalenbach Business Review* 64(2): 281-307.

Köhler, A. G. and J. C. Theis. 2018. Die Entwicklung der Abschlussprüferhonorare in Deutschland. *Betriebswirtschaftliche Forschung und Praxis* 70(4): 390-415.

Knechel, W. R. 2009. *Audit lessons from the economic crisis: Rethinking audit quality*. Inaugural lecture at Maastricht University.

Knechel, W. R., G. V. Krishnan, M. Pevzner, L. B. Shefchik, and U. K. Velury. 2013. Audit quality: Insights from the academic literature. *Auditing: A Journal of Practice & Theory* 32(1): 385-421.

Krauß, P., P. Pronobis, and H. Zülch. 2015. Abnormal audit fees and audit quality: initial evidence from the German audit market. *Journal of Business Economics* 85(1): 45-84.

Krauß, P. and H. Zülch. 2013. Non-audit services and audit quality: Blessing or curse? *The Journal of Applied Business Research* 29(2): 305-325.

Lange, S. 1994. *Die Kompatibilität von Abschlussprüfung und Beratung*. Frankfurt am Main: Europäischer Verlag der Wissenschaften.

Larcker, D. F. and S. A. Richardson. 2004. Fees paid to audit firms, accrual choices, and corporate governance. *Journal of Accounting Research* 42(3): 625-658.

Larcker, D. F. and T. O. Rusticus. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* 49(3): 186-205.

Laschewski, C., M. Möller, and M. Risse. 2014. Eine empirische Analyse der Folgekosten des Enforcements der Rechnungslegung durch die Deutsche Prüfstelle für Rechnungslegung. *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung (KoR)* 14(6): 307-312.

Lehrbass, F. and N. Scheipers. 2017. Determinanten der Höhe von Abschlussprüfungshonoraren. Eine empirische Untersuchung von Unternehmen im Prime Standard. *Die Wirtschaftsprüfung* 70(24): 1437-1443.

Leidner, J.J. and H. Lenz. 2013. Kreditinstitute und die Konzentration des deutschen Marktes für Abschlussprüferleistungen. *Die Betriebswirtschaft* 73(5): 379-400.

Lenz, H. 2016. Honorare für Abschlussprüfungs- und Nichtprüfungsleistungen bei Unternehmen von öffentlichem Interesse. *Der Betrieb* 44: 2555-2560.

Lenz, H., M. Möller, and B. Höhn. 2006. Offenlegung der Honorare für Abschlussprüferleistungen im Geschäftsjahr 2005 bei DAX-Unternehmen. *Betriebs-Berater* 61(33): 1787-1793.

Lopatta, K., T. Kaspereit, F. Canitz, and T. Maas. 2015. Abschlussprüfung und Honorarabhängigkeit - Eine empirische Untersuchung der Auswirkungen auf die Unabhängigkeit des Abschlussprüfers. *Betriebswirtschaftliche Forschung und Praxis* 67(5): 562-591.

Loy, T. and S. Steuer. 2015. Vertrauen ist gut, Enforcement ist besser? Eine Bilanz nach zehn Jahren DPR (Teil 2). *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung (KoR)* 15(11): 548-557.

Luik, H. 1976. Ist ein obligatorischer Prüferwechsel für Aktiengesellschaften sinnvoll? *Betriebs-Berater* 31: 237-239.

MacCarthy, J. 2017. Using Altman Z-score and Beneish M-score models to detect financial fraud and corporate failure: A case study of Enron corporation. *International Journal of Finance and Accounting* 6(6): 159-166.

Markelevich, A. and R. L. Rosner. 2013. Auditor fees and fraud firms. *Contemporary Accounting Research* 30(4): 1590-1625.

Marten, K.-U., R. Quick, and K. Ruhnke. 2015. *Wirtschaftsprüfung. Grundlagen des betriebswirtschaftlichen Prüfungswesens nach nationalen und internationalen Normen*. Stuttgart: Schäffer-Poeschel.

McNichols, M. F. 2000. Research design issues in earnings management studies. *Journal of Accounting and Public Policy* 19(4-5): 313-345.

Mehta, A. and G. Bhavani. 2017. Application of forensic tools to detect fraud: The case of Toshiba. *Journal of Forensic and Investigative Accounting* 9(1): 692-710.

- Meuwissen, R. and R. Quick. 2009. Abschlussprüfung und Beratung - Eine experimentelle Analyse der Auswirkungen auf Unabhängigkeitswahrnehmungen deutscher Aufsichtsräte. *Schmalenbachs Zeitschrift für Betriebswirtschaftliche Forschung* 61(4): 382-415.
- Molls, F. 2013. *Der Einfluss von Mandatsdauer, Rotation und Nichtprüfungshonoraren auf die Qualität der Abschlussprüfung. Eine empirische Untersuchung*. Hamburg: Verlag Dr. Kovac.
- Ostrowski, M. and B. H. Söder. 1999. Der Einfluss von Beratungsaufträgen auf die Unabhängigkeit des Jahresabschlussprüfers. *Betriebswirtschaftliche Forschung und Praxis* 5: 554-564.
- Palmrose, Z.-V. 1986. The effect of nonaudit services on the pricing of audit services: further evidence. *Journal of Accounting Research* 24(2): 405-411.
- Paolone, F. and C. Magazzino. 2014. Earnings manipulation among the main industrial sectors. Evidence from Italy. *Economia Aziendale Online* 5(4): 253-261.
- Pasch, L. 2017. What drives misstatements in financial statements? Evidence from Germany. Working paper. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3062805.
- Paul, M. 2011. Enforcement der Rechnungslegung kapitalmarktorientierter Unternehmen aus der Sicht der Rechtsprechung. *Die Wirtschaftsprüfung* 64(1): 11-16.
- Pott, C., T. Schröder, and L. Weckelmann. 2014. Externe Rotation und Verbot von Nichtprüfungsleistungen. Lösung zur Verbesserung der Prüfungsqualität? *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung* 14(10): 499-504.
- Powers, E. A. 2005. Interpreting logit regressions with interaction terms: An application to the management turnover literature. *Journal of Corporate Finance* 11(3): 504-522.
- PWC and DAI. 2013. Bilanzkontrolle 2013. Available at: https://www.dai.de/files/dai_usercontent/dokumente/studien/2013-10-15%20DAI-PwC-Studie%20Bilanzkontrolle%202013.pdf.
- Qandil, J. S. 2014. *Wahrnehmung der Qualität der Abschlussprüfung. Eine theoretische und empirische Analyse für den deutschen Kapitalmarkt*. Wiesbaden: Springer Fachmedien.
- Quick, R. and M. Sattler. 2011a. Das Erfordernis der Umsatzunabhängigkeit und die Konzentration auf dem deutschen Markt für Abschlussprüferleistungen. *Zeitschrift für Betriebswirtschaft* 81(1): 61-98.
- Quick, R. and M. Sattler. 2011b. Beeinträchtigen Beratungsleistungen die Urteilsfreiheit des Abschlussprüfers? Zum Einfluss von Beratungshonoraren auf diskretionäre Periodenabgrenzungen. *Schmalenbachs Zeitschrift für Betriebswirtschaftliche Forschung* 63(4): 310-343.
- Quick, R., F. Schmidt, and D. Simons. 2016. Wo liegen Schwachstellen der Joint-Audit-Forschung? *Die Wirtschaftsprüfung* 69(4): 195-200.

Quick, R. and B. Warming-Rasmussen. 2009. Auditor independence and the provision of non-audit services: Perceptions by German investors. *International Journal of Auditing* 13(2): 141-162.

Quick, R. and B. Warming-Rasmussen. 2015. An experimental analysis of the effects of non-audit services on auditor independence in appearance in the European Union: Evidence from Germany. *Journal of International Financial Management & Accounting* 26(2): 150-187.

Quick, R. and D. Wiemann. 2011. Zum Einfluss der Mandatsdauer des Abschlussprüfers auf die Prüfungsqualität. *Zeitschrift für Betriebswirtschaft* 81(9): 915-943.

Quick, R. and D. Wiemann. 2013. Externe Pflichtrotation des Abschlussprüfers, Prüfungsmarkt und Prüfungsqualität. Eine Diskussion der EU-Vorschläge im Lichte von Forschungsbefunden. *Zeitschrift für Corporate Governance* 2: 77-85.

Ratzinger-Sakel, N. V. 2013. Auditor fees and auditor independence - Evidence from going concern reporting decisions in Germany. *Auditing: A Journal of Practice & Theory* 32(4): 129-168.

Ratzinger-Sakel, N. V. and M. W. Schönberger. 2015. Restricting non-audit services in Europe—The potential (lack of) impact of a blacklist and a fee cap on auditor independence and audit quality. *Accounting in Europe* 12(1): 61-86.

Repousis, S. 2016. Using Beneish model to detect corporate financial statement fraud in Greece. *Journal of Financial Crime* 23(4): 1063-1073.

Ronen, J. and V. Yaari. 2008. *Earnings management. Emerging insights in theory, practice, and research*. Boston, MA: Springer US.

Rossmann, J. and T. Bleicher. 2017. Nutzen einer Abschlussprüfung - Herausforderungen und Spannungsfelder im internationalen Kontext. In W. Funk and J. Rossmann. (eds.). *Internationale Rechnungslegung und Internationales Controlling: Herausforderungen – Handlungsfelder - Erfolgspotenziale* (pp. 213-234). Wiesbaden: Springer Gabler.

Roxas, M. L. 2011. Financial statement fraud detection using ratio and digital analysis. *Journal of Leadership, Accountability, and Ethics* 8(4): 56-66.

Sattler, M. 2011. *Vereinbarkeit von Abschlussprüfung und Beratung*. Wiesbaden: Springer Fachmedien.

Sharma, D. S. and J. Sidhu. 2001. Professionalism vs commercialism: The association between non-audit services (NAS) and audit independence. *Journal of Business Finance & Accounting* 28(5-6): 595-629.

Scheffler, E. 2006. Aufgaben und erste Erfahrungen des Enforcements. *Zeitschrift für internationale Rechnungslegung* 1(1): 13-21.

Schmidt, S., and C. Almeling. 2018. HGB § 317 Gegenstand und Umfang der Prüfung, in Beck'scher Bilanz-Kommentar, 11th edition. Munich.

Schmidt, S. and T. Nagel. 2018. HGB § 319 Auswahl der Abschlussprüfer und Ausschlussgründe, in Beck'scher Bilanz-Kommentar, 11th edition. Munich.

Schmidt, S. and T. Nagel. 2018. HGB § 319a Besondere Ausschlussgründe bei Unternehmen von öffentlichem Interesse, in Beck'scher Bilanz-Kommentar, 11th edition. Munich.

Sheskin, D. J. 2000. *Handbook of Parametric and Nonparametric Statistical Procedures*. 2nd edition. Boca Raton, FL: Chapman & Hall/CRC.

Shipman, J. E., Q. T. Swanquist, and R. L. Whited. 2017. Propensity score matching in accounting research. *The Accounting Review* 92(1): 213-244.

Sinha, P. 2008. Auditor independence and the likelihood of a GAAP violation. Working paper. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1139143.

Simnett, R., E. Carson, and A. Vanstraelen. 2016. International archival auditing and assurance research: trends, methodological issues, and opportunities. *Auditing: A Journal of Practice and Theory* 35(3): 1-32.

Simon-Heckroth, E. and M. Lüdders. 2017. Anhangangaben über das Abschlussprüferhonorar. IDW RS HFA 36 n.F. im Kontext der Begrenzung der Erbringung von Nichtprüfungsleistungen nach der EU-VO. *Die Wirtschaftsprüfung* 70(5): 248-256.

Simunic, D. A. 1980. The pricing of audit services: Theory and evidence. *Journal of Accounting Research* 18(1): 161-190.

Simunic, D. A. 1984. Auditing, consulting and auditor independence. *Journal of accounting research* 22(2): 679-702.

Strohmeier, M. 2013. Enforcement of accounting standards in Germany: The pre-and post-misstatement development of censured firms. Working paper. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2288622.

Strohmeier, M. 2014. Enforcement releases, firm characteristics, and earnings quality: Insights from Germany's two-tiered Enforcement system. *Journal of International Financial Management & Accounting* 25(3): 271-304.

Stuart, E. A. 2010. Matching methods for causal inference: A review and a look forward. *Statistical Science* 25(1): 1-21.

Stuart, I., Y. C. Shin, D. P. Cram, and V. Karan. 2013. Review of choice-based, matched, and other stratified sample studies in auditing research. *Journal of Accounting Literature* 32(1): 88-113.

Sultana, A. and C. Willeke. 2005. Die neuen Anhangangaben durch das Bilanzrechtsreformgesetz – Darstellung und Erläuterung von § 285 Satz 1 Nr. 17 bis 19 HGB. *StuB – Unternehmensteuern und Bilanzen* 22: 951-970.

Tarjo, N. H. 2015. Application of Beneish M-Score models and data mining to detect financial fraud. *Procedia-Social and Behavioral Sciences* 7(211): 924-930.

Tebben, T. 2011. *Vergütungsanreize und opportunistische Bilanzpolitik. Eine empirische Analyse der Rolle von Aufsichtsrat und Abschlussprüfer*. Wiesbaden: Gabler Verlag.

Tepalagul, N. and L. Lin. 2015. Auditor independence and audit quality: A literature review. *Journal of Accounting, Auditing and Finance* 30(1): 101-121.

Thormann, B. and D. Barth. 2018. Aktuelles aus der DPR Enforcement Prüfung. In P. Velte, S. Müller, S. C. Weber, R. Sassen, and A. Mammen (eds). *Rechnungslegung, Steuern, Corporate Governance, Wirtschaftsprüfung und Controlling* (pp. 473-489). Wiesbaden: Springer Fachmedien.

Thormann, B. and I. Zempel. 2017. Enforcement der Rechnungslegung in Deutschland. In J. Baetge, P. Wollmert, H. J. Kirsch, P. Oser, and S. Bischof (eds.). *Rechnungslegung nach IFRS: Kommentar auf der Grundlage des deutschen Bilanzrechts*. Stuttgart: Schäffer-Poeschel.

Toebe, M., A. Herberg, and M. Schlüter. 2013. Untersuchung der Fehlerbekanntmachungen nach § 37q Abs. 2 WpHG von 2006 bis 2013. Ergebnisse und Schlussfolgerungen. *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung (KoR)* 13(5): 256-260.

Tritschler, J. 2013. *Audit quality. Association between published reporting errors and audit firm characteristics*. Wiesbaden: Springer Fachmedien.

Velte, P. 2016. Der Regierungsentwurf für ein Abschlussprüfungsreformgesetz (AREGRegE). Künftiges Zusammenspiel von Aufsichtsrat und Abschlussprüfer. *Die Wirtschaftsprüfung* 69(3): 125-131.

Velte, P. and T. Loy. 2018. The impact of auditor rotation, audit firm rotation and non-audit services on earnings quality, audit quality and investor perceptions: A literature review. *Journal of Governance and Regulation* 7(2): 74-90.

Van Liempd, D., R. Quick, and B. Warming-Rasmussen. 2019. Auditor-provided nonaudit services: Post-EU-regulation evidence from Denmark. *International Journal of Auditing* 23(1): 1-19.

Wagenhofer, A. and R. Ewert. 2015. *Externe Unternehmensrechnung*. 3rd edition. Berlin Heidelberg: Springer-Verlag.

Wagner, W. 2011. *Modelle zur Aufdeckung von Earnings Management: Eine kritische Analyse und Optimierungsansätze*. Hamburg: Diplomica Verlag.

Wiedman, C. 2002. The power of auditors: auditors and regulators can restore faith in financial reporting systems. *CA Magazine* 135(10): 39-44.

Wooldridge, J. M. 2016. *Introductory econometrics: a modern approach*. 6th edition. Boston, MA: Cengage Learning.

WPK. 2009. Transparenzbericht gemäß § 55c WPO – Bestandsaufnahme und Gestaltungshinweise. *WPK Magazin* 1: 4-12.

Zimmermann, R.-C. 2006. Gestaltungsspielräume bei Veröffentlichung von Abschlussprüferhonoraren im Rahmen des BilReG. *Zeitschrift für internationale und kapitalmarktorientierte Rechnungslegung (KoR)* 6(4): 273-276.

Zimmermann, R.-C. 2008. *Abschlussprüfer und Bilanzpolitik der Mandanten*. Wiesbaden: Gabler.

Zmijewski, M. E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research* 22: 59-82.

Zülch, H. 2005. Das deutsche Enforcement-Modell des Bilanzkontrollgesetzes – Ausgestaltung und Implikationen für Rechnungslegung und Abschlussprüfung. *Steuern und Bilanzen* 1: 1-9.

Zülch, H. and M. Hendler. 2017. *Bilanzierung nach IFRS*. Weinheim: Wiley-VCH Verlag & Co. KGaA.

Zwernemann, J. 2015. *Forensic Services*. Wiesbaden: Springer Fachmedien.