

ESG-Performance in der deutschen Versicherungswirtschaft

Tim Dahme und Frederick Schuh

Zusammenfassung

Im Rahmen dieser Arbeit wurde untersucht, ob ausgewählte Nachhaltigkeitsratings dazu geeignet sind, die ESG-Performance in der deutschen Versicherungsbranche angemessen zu erfassen und abzubilden. Die Ergebnisse zeigen, dass die betrachteten ESG-Ratings aufgrund fehlender signifikanter statistischer Zusammenhänge sowie großer Abweichungen in den Scores und Bewertungsmaßstäben nur eingeschränkt geeignet sind, die ESG-Performance zu messen. Besonders hervorzuheben sind die zum Teil erheblichen absoluten Unterschiede zwischen den Rating-Scores eines einzelnen Unternehmens, die zu grundlegend unterschiedlichen Einschätzungen der Nachhaltigkeit desselben Unternehmens führen können.

Abstract

This study investigated whether selected sustainability ratings are suitable for adequately capturing and depicting ESG performance in the German insurance industry. The results show that the ESG ratings examined are only suitable for measuring ESG performance to a limited extent due to a lack of significant statistical correlations and large deviations in the scores and assessment criteria. Particularly noteworthy are the sometimes considerable absolute differences between the rating scores of an individual com-

Tim Dahme
Universität zu Köln
Seminar für ABWL, Risikomanagement und Versicherungslehre
Kerpener Straße 30
50937 Köln
E-Mail: tim.dahme2@gmail.com

Frederick Schuh (corresponding author)
Universität zu Köln
Seminar für ABWL, Risikomanagement und Versicherungslehre
Kerpener Straße 30
50937 Köln
<https://orcid.org/0000-0001-9042-2458>
E-Mail: frederick.schuh@wiso.uni-koeln.de

pany, which can lead to fundamentally different assessments of the sustainability of the same company.

JEL classification: G22, M14, Q01

Keywords: ESG-Performance, Nachhaltigkeitsratings, Versicherung, ESG-Scores

1. Einleitung

Einhergehend mit der stetig zunehmenden Bedeutung der Nachhaltigkeit in allen Gesellschaftsbereichen rückt verstärkt das Agieren von Unternehmen unter entsprechenden Gesichtspunkten in den Fokus der öffentlichen Debatte. So sind Unternehmen häufig aufgefordert darzulegen, wie sie sich im Hinblick auf spezifische Themenbereiche, aber auch ganzheitlich im Sinne der Nachhaltigkeit positionieren. Ziel ist es aus Sicht der jeweiligen Anspruchsgruppen (z.B. Privatpersonen, öffentliche Einrichtungen, Investoren) in der Regel die Nachhaltigkeitsperformance eines Unternehmens oder einer ganzen Branche zu bewerten. Dabei hat sich insbesondere auf der wirtschaftlichen Ebene die Unterteilung der Nachhaltigkeit in die Kategorien „Environmental“, „Social“ und „Governance“ (ESG) etabliert. Diese Bereiche umfassen mitunter eine Vielzahl von weiteren Themenfeldern und sind in der jeweiligen Auslegung oftmals subjektiver Natur. Erste Schritte der Standardisierung und Regulierung zur Offenlegung von nachhaltigkeitsbezogenen Informationen von Unternehmen finden derzeit unter anderem mit der Einführung der *Corporate Sustainability Reporting Directive* durch die *Europäische Union* statt (Europäisches Parlament 2022). Um unter diesen Voraussetzungen die Nachhaltigkeitsperformance zu messen, sind strukturierte und fundierte Bewertungsansätze notwendig. Da es den einzelnen Interessensgruppen in der Regel kaum möglich ist entsprechende benötigte Ressourcen aufzubringen, haben sich über die Zeit verschiedene Anbieter auf die Entwicklung und Anwendung solcher Bewertungsansätze zur Erfassung der Nachhaltigkeit einzelner Unternehmen spezialisiert. Durch die Nutzung der ESG-Scores speziell im Rahmen von Investitionsentscheidungen, ergibt sich eine unmittelbare Relevanz für viele Branchen und Unternehmen. Dies gilt im Besonderen für die Versicherungsbranche, welche unter anderem im Rahmen der Verwaltung und Kapitalanlage von Versicherungsbeiträgen signifikante Kapitalströme bewegt. Somit ergibt sich einerseits eine große Bedeutung in Bezug auf entsprechende Anlageentscheidungen, andererseits aber auch eine Verbindung zu Versicherungsnehmern, welche ihrerseits Nachhaltigkeitsaspekte in die Entscheidung für oder gegen eine einzelne Versicherung fließen lassen können.

Aufgrund der gestiegenen Bedeutung der ESG-Scores, sowie der Herausforderung im Rahmen der quantitativen Erfassung eines subjektiven Themas wie der Nachhaltigkeit, untersucht eine Reihe von Forschungsarbeiten verschiedene As-

pekte im Rahmen der Nachhaltigkeitsratings. Innerhalb dieser wissenschaftlichen Literatur hat sich ein Strang herausgebildet, welcher sich mit der Korrelation der Ratings untereinander und entsprechenden Schlussfolgerungen auseinandersetzt. Grundsätzlich soll dabei herausgestellt werden, ob die verschiedenen Ratings eine ähnliche Sicht auf die Nachhaltigkeit der beurteilten Unternehmen besitzen oder sich an dieser Stelle Unterschiede zeigen, welche die Aussagekraft beeinträchtigen. Berg et al. (2022) dokumentieren in ihrer Untersuchung geringe paarweise Korrelationen unter anderem für die Ratings von Sustainalytics, S&P Global, Refinitiv und MSCI mit einem Gesamtdurchschnitt von 0,54 (Berg, Kölbel, and Rigobon 2022). Die Autoren identifizieren mehrere Quellen, aus welchen die geringen statistischen Zusammenhänge resultieren. Darüber hinaus berechnen sie für klassische Kreditratings als Vergleichswert eine Korrelation von 0,99 (Berg, Kölbel, and Rigobon 2022). Im Rahmen der Analyse von Christensen et al. (2022) bestätigen diese zunächst das Vorliegen signifikanter Differenzen zwischen den Ratings und führen im Weiteren aus, dass sich diese im Laufe der Zeit verstärken (Christensen et al. 2020). Außerdem zeigen die Autoren, dass Abweichungen zwischen den Rating-Scores insbesondere im Falle niedriger Scores stark ausgeprägt sind und in positivem Zusammenhang mit einem höheren Niveau an Veröffentlichung von Daten aus den Bereichen „Umwelt“ und „Soziales“ stehen. Serafeim und Yoon (2022) demonstrieren in ihrer Arbeit zusätzlich, dass grundsätzlich ein stark positiver Zusammenhang zwischen ESG-Ratings und zukünftigen ESG-News besteht, woraus sich zunächst ableitet, dass Nachhaltigkeitsratings eine Abbildung der Realität mit Bezug auf die Nachhaltigkeit der Unternehmen vornehmen. Dennoch nehmen auch sie Bezug auf eine fehlende Übereinstimmung zwischen den Ratings und weisen darauf hin, dass insbesondere bei Betrachtung der Quelle der Messbarkeit gemäß Berg et al. (2022) die Aussagekraft der Ratings mit Blick auf zukünftige ESG-News abnimmt (Serafeim and Yoon 2021). Gibson Brandon et al. (2021) stellen für eine Reihe von Nachhaltigkeitsratings eine durchschnittliche Korrelation von 0,45 fest und weisen darauf hin, dass für bestimmte Kombinationen von Ratings vergleichsweise hohe paarweise Korrelationen beobachtet werden können (Gibson Brandon, Krueger, and Schmidt 2021). Die Autoren nehmen des Weiteren eine Analyse dahingehend vor, ob die Unterschiede in den ESG-Ratings Auswirkungen auf die Aktienrendite haben. Dabei stellen sie einen positiven Zusammenhang fest und folgern, dass daraus ein signifikantes Risiko für die finanzielle Bewertung von Unternehmen entsteht (Gibson Brandon, Krueger, and Schmidt 2021). Während sich die soeben genannten Untersuchungen auf die jüngste Vergangenheit beziehen, haben Chatterji et al. (2016) bereits gezeigt, dass fehlende statistische Zusammenhänge zwischen verschiedenen ESG-Ratings zu beobachten sind und diese daher einer kritischen Betrachtung bedürfen (Chatterji et al. 2016). Daran anknüpfend beobachten Semenova und Hassel (2015) für eine vergleichsweise sehr kleine Gruppe von drei Ratings, wel-

che in der Zwischenzeit durch Übernahmen in anderen Anbietern aufgegangen sind, partiell hohe Korrelationen (Semenova and Hassel 2015).

Unsere Analyse untersucht, ob es einem unabhängigen Externen anhand der Ratings möglich ist, ein Unternehmen im Hinblick auf dessen Nachhaltigkeit angemessen zu beurteilen. Dazu soll mit Korrelationsmaßen ermittelt werden, in welchem Verhältnis sich die Ratings zueinander befinden. Eine Beantwortung der Fragen erfolgt auf Basis einer rein externen Perspektive, d.h. dass ausschließlich öffentlich verfügbare Informationen verwendet werden. Entsprechend dem gesetzten Rahmen gilt, dass Nachhaltigkeit in dieser Arbeit analog zu den ESG-Kriterien aufgegriffen und dabei ganzheitlich und nicht mit einer Unterscheidung zwischen den drei Bereichen betrachtet wird.

Im Folgenden werden in Kapitel 2 zunächst relevante ESG-Ratings und -Scores dargelegt bevor Kapitel 3 das methodische Vorgehen darstellt. Die Analyse und Darstellung der Ergebnisse erfolgen in Kapitel 4 zunächst anhand verschiedener statistischer Maße. In Kapitel 5 werden die Analyseergebnisse diskutiert und die Arbeit in Kapitel 6 im Rahmen des Fazits zusammengefasst.

2. Datengrundlage und Harmonisierung

Mit Blick auf eine Eingrenzung des Umfangs des Vorgehens und einer gleichzeitigen ausreichenden Abdeckung des Gesamtmarktes, wird die Anzahl der in die Untersuchung einzubeziehenden Versicherungsunternehmen auf die 50 größten (auch: Top 50) begrenzt. Hierbei werden noch vor weiteren Schritten die öffentlich-rechtlichen Versicherer aufgrund ihrer speziellen und ggf. nur eingeschränkt mit privatwirtschaftlichen Unternehmen vergleichbaren Natur ausgeschlossen. die finale Auswahl an Unternehmen nach der Bereinigung repräsentiert circa 86 % des deutschen Gesamtmarktes. Die verbleibenden Versicherungsunternehmen werden im nächsten Schritt zunächst auf die jeweils verfügbaren Berichte hin untersucht. Um auch hierbei einen ausgewogenen Umfang zu wahren, werden die folgenden Berichtsarten berücksichtigt: Gesonderte Nachhaltigkeitsberichte (engl.: Sustainability report; auch andere Bezeichnungen wie z. B. CSR-Bericht möglich), sogenannte Nicht-Finanzielle Erklärungen und Geschäftsberichte mit integrierter Nachhaltigkeitsberichtserstattung.

Darauf aufbauend erfolgt die Auswertung der in den Berichten enthaltenen Nennungen von Nachhaltigkeitsratings. In einem zweiten Schritt wird in jedem Fall der Internetauftritt des jeweiligen Unternehmens auf Nennungen hin durchsucht. Anschließend werden sämtliche Nennungen in einer Auflistung zusammengeführt. Es gilt zu beachten, dass jedes Rating nur einmal pro Unternehmen gelistet wird – unabhängig davon, ob es in mehreren der Quellen identifiziert wird. Wie in Tabelle 1 erkenntlich, verbleiben neun ESG-Ratings in der finalen Gesamtheit. Grund ist zum einen die zwingend notwendige öffentliche

Verfügbarkeit der Rating-Scores für eine weitere Nutzung im Rahmen der statistischen Analyse. Zum anderen wird im Einzelfall betrachtet, ob eine ausreichende Anzahl an Scores der deutschen Versicherungsunternehmen im Einzelnen und in Überschneidung mit anderen Ratings vorliegt.

Tabelle 1
Übersicht über die final ausgewählten ESG-Ratings¹

Rating ID	Name
#01	Refinitiv ESG Score
#02	ISS ESG Rating
#03	Sustainalytics ESG Rating
#04	Zielke ESG-Unternehmensranking
#05	S&P CSA
#06	MSCI ESG Rating
#07	ShareAction Report Score
#08	IVFP Kompetenzrating Nachhaltigkeit
#09	Franke und Bornberg ESG-Rating

Zur Erhebung der Scores dienen im Wesentlichen zwei Quellen. Einige Ratings veröffentlichen jährliche Rankings oder Berichte, in welchem alle beurteilten Unternehmen mit ihrem jeweiligen Ergebnis aufgeführt und die benötigten Daten der hier relevanten Unternehmen entnommen werden können.² Andere Ratings nutzen website-basierte Datenbanken mit einer entsprechenden Suchfunktion.³ Bei einigen Versicherungsunternehmen sind Daten nur auf Ebene der (teilweise nicht deutschen) Muttergesellschaft vorhanden. Diese Daten werden explizit nicht ausgeschlossen, um eine grundsätzliche Annäherung an die entsprechende deutsche Unternehmenseinheit mit einzubeziehen.⁴ Ein Min-

¹ Aus Gründen der Übersichtlichkeit wird das ESG-Unternehmensranking von Morgen & Morgen und Zielke in dieser Arbeit als „Zielke ESG-Unternehmensranking“ bzw. dem „ESG-Score“ von Zielke geführt. Diese treten innerhalb der Kooperation der Unternehmen federführend bei der Präsentation des Rankings auf.

² Mit Ausnahme des ShareAction Report Score, welcher einem einzelnen Ranking aus dem Jahre 2021 entstammt.

³ Die Erhebung der Daten erfolgte im Zeitraum 14.–16. April 2023.

⁴ Spezieller Fall der ERGO Versicherung: Die Scores sind überwiegend nur auf Ebene der Muttergesellschaft (Münchener Rück) verfügbar. In einem Einzelfall stehen beide

destmaß an Übereinstimmung in der Nachhaltigkeitsperformance zwischen Gesamtunternehmen und Tochtergesellschaft wird hierbei entsprechend angenommen. Ein analoges Vorgehen wird in solchen Fällen angewandt, in denen Scores nur einzeln auf Ebene der Kranken- und/oder Lebensversicherungseinheit verfügbar sind. Hier wird das einzelne Rating im Ermangelung übergreifender Daten stellvertretend für das gesamte Unternehmen verwendet oder ein Durchschnitt gebildet.

Tabelle 2
Übersicht der Unterschiede der verschiedenen Ratingskalen

Rating ID	Rating	Original Skala schlecht/gut	Richtung der Skala <i>Höhere Punkte = besser/schlechter</i>
#01	Refinitiv ESG Score	0 / 100	besser
#02	ISS ESG Rating	D- / A+	besser
#03	Sustainalytics ESG Rating	49,99 / 0	schlechter
#04	Zielke ESG-Unternehmensranking	-4,67 / 5,25	besser
#05	S&P CSA	0 / 100	besser
#06	MSCI ESG Rating	CCC / AAA	besser
#07	Share Action Report Score	0% / 100%	besser
#08	IVFP Kompetenzrating Nachhaltigkeit	Gut / Sehr Gut / Exzellent	schlechter
#09	Franke und Bornberg ESG-Rating	0% / 100%	besser

Quelle: Eigene Darstellung in Anlehnung an Gibson Brandon et al. (2021).

Um eine angemessene Vergleichbarkeit der Ergebnisse aus den jeweiligen Ratings herzustellen und anhand dieser in der Analyse Rückschlüsse ziehen zu können, ist es notwendig zunächst eine Harmonisierung der Rating-Skalen und darauf basierend der Datenbasis vorzunehmen. Ziel ist es, die Daten in einer metrischen Form und auf einer gleichen Skala abzubilden. Eine solche Skala

Scores zur Verfügung, hier wird aus Gründen der höchstmöglichen Einheitlichkeit auf den Score der Münchener Rück zurückgegriffen. In einem weiteren Fall ist nur der ERGO-Score verfügbar – dieser wird aus Gründen der höchstmöglichen Vollständigkeit verwendet.

wird analog zum Vorgehen von Christensen et al. (2022) als 0 bis 100 definiert, wobei ein höherer Wert einer besseren ESG-Performance entspricht (Christensen et al. 2020). Anhand von Tabelle 2 wird ersichtlich, dass das Skalenniveau der Ratings *ISS*, *MSCI* und *IVFP* ordinal ist. Konkret bedeutet dies, dass es zwar möglich ist, Rating Ergebnisse in eine Rangfolge zu bringen, aus welcher jedoch nicht die Größe der Abstände hervorgeht. Zudem liegen die öffentlich verfügbaren Rohdaten in Form von Buchstaben oder qualitativen Klassen vor. Um eine Auswertung hinsichtlich der Korrelation zwischen den Ratings sowie weitere Vergleiche vorzunehmen, werden die Daten entsprechend in ein zahlenbasiertes Format umgewandelt. Für die Ratings von *ISS* und *MSCI* wird die Skalenbreite gleichmäßig auf die jeweilige Anzahl der Klassen verteilt, wodurch sich für jede Klasse eine Ober- und Untergrenze ergibt⁵. Basierend darauf wird für die weitere Verwendung der jeweilige Klassenmittelpunkt bestimmt. So führt beispielsweise der Rating-Score „AA“ von *MSCI* zu einer Klassenunter- und -obergrenze von 71,43 und 85,71 mit einem Mittelpunkt von 78,57. In der Folge wird jedes Unternehmen, welches den Score „AA“ von *MSCI* erhält, mit diesem Mittelpunkt in die weitere Auswertung einbezogen. Da keine genaueren Informationen gegeben sind, an welcher Stelle sich der Score innerhalb einer Klasse befindet, wird der jeweilige Mittelpunkt verwendet. Das Rating des *IVFP* gibt die jeweiligen quantitativen Grenzen der drei verwendeten Klassen an. Hier ist zu beachten, dass einerseits die Breite der Klassen nicht gleichmäßig ist und andererseits eine höhere Punktzahl zu einem schlechteren Rating führt.

Es verbleiben zwei weitere Nachhaltigkeitsratings (*Sustainalytics*, *Zielke*), welche bereits in metrischer Form vorliegen, allerdings eine jeweils andere Punkteskala verwenden. In beiden Fällen gilt zur Umwandlung der Datenpunkte eine Formel, welche die einzelnen Scores in das Verhältnis zur jeweiligen Skalenbandbreite setzt und anschließend mit der neuen, benötigten Bandbreite von 100 multipliziert.⁶ Im Falle des ESG-Scores von *Sustainalytics* wird ebenfalls mit einbezogen, dass die Rohdaten analog zu *IVFP* in umgekehrter Richtung dargestellt sind.

⁵ Die Methodologie des *MSCI* ESG Ratings beinhaltet neben den Bezeichnungen der Klassen mit Buchstaben auch quantitative Ober- und Untergrenzen, welche auf einer Skala von 0 – 10 ebenfalls gleichmäßig verteilt sind und den hier bestimmten Werten reduziert um den Faktor zehn entsprechen.

⁶ *Sustainalytics* ordnet die einzelnen Scores sogenannten Risikoklassen zu, welche hier nicht genutzt werden, da die granulare und punktgenaue Form der Daten notwendig ist. Aus den Klassen geht keine Begrenzung der möglichen Punktzahl nach oben hervor, weshalb für die höchste, angegebene Klasse die gleiche Breite wie im Falle der anderen Klassen angenommen wird. Hieraus ergibt sich die Gesamtobergrenze von 49.99.

3. Methodik

Die im ersten Teil der Analyse verwendeten statistischen Maße beinhalten zwei Varianten von Korrelationskoeffizienten – *Bravais-Pearson* und *Spearman* – sowie die durchschnittliche absolute Abweichung der Scores. Die Auswahl geeigneter statistischer Mittel erfolgt dabei auf Basis des ausgewählten Forschungsstandes (siehe Kapitel 1). Unter anderem wird in den Beiträgen von Berg et al. (2022), Christensen et al. (2022) sowie Serafeim und Yoon (2022) anhand von verschiedenen Maßen ermittelt, ob und wie stark die Scores der Ratings bei gleichen Unternehmen voneinander abweichen. Eine solche Betrachtung erweist sich zur Beantwortung der Forschungsfragen als unerlässlich, da der Nachweis deutlicher Unterschiede in den Ratingergebnissen Rückschlüsse auf eine ggf. mangelnde Eignung zur Erfassung der ESG-Performance in der deutschen Versicherungsbranche zulässt. Der Hintergrund dafür ist im Wesentlichen die in solchen Fällen potentiell fundamental voneinander abweichenden Aussagen über die Nachhaltigkeit eines Unternehmens und die damit einhergehende geschmälerte Aussagekraft. Im Umkehrschluss bedeutet dies, dass einzelne Ratings keine angemessene Bewertungsgrundlage für unbeteiligte Externe darstellen, um letztlich die ESG-Performance eines Unternehmens beurteilen zu können. Darüber lässt sich ebenfalls die Stärke der Unterschiede bestimmen, was es ermöglicht, eine Aussage darüber zu treffen, welche der Ratings im Einzelnen besser geeignet sind als andere.

Als erste Komponente der statistischen Analyse wird der Korrelationskoeffizient nach Bravais-Pearson betrachtet, welcher unter anderem ebenfalls in den Arbeiten von Berg et al. (2022), Gibson Brandon et al. (2021) und Chatterij et al. (2016) genutzt wird. Dieser bestimmt paarweise den linearen Zusammenhang zwischen zwei Variablen – im Folgenden jeweils die einzelnen Kombinationen der im Vorhinein ausgewählten Ratings. Die Formel für den Bravais-Pearson-Korrelationskoeffizient lautet:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Dabei gilt für den Korrelationskoeffizienten per Definition folgender möglicher Wertebereich:

$$-1 \leq r_{xy} \leq 1$$

Aus den notwendigen Voraussetzungen für den Koeffizienten ergeben sich in Verbindung mit dem hier vorliegenden Datenset weitere grundlegende Einschränkungen. Im Rahmen der Betrachtung der Skalen wurde bereits dargelegt,

dass einige Ratings die Scores in ordinaler Datenform zur Verfügung stellen. Für eine korrekte Ermittlung des Bravais-Pearson-Korrelationskoeffizienten werden allerdings metrische, intervallskalierte Daten benötigt. Diese Bedingung wird durch die bereits vorgenommene Umwandlung zwar erreicht, allerdings verbleibt eine daraus hervorgehende Ungenauigkeit, welche aus einer methodischen Perspektive zu einer Einschränkung bezüglich der Interpretation der Ergebnisse führen kann.

In Ergänzung zum Bravais-Pearson-Korrelationskoeffizienten wird in Folge dessen der Korrelationskoeffizient nach Spearman (auch Rangkorrelationskoeffizient genannt) herangezogen, um die aus statistischer und methodischer Sicht nur eingeschränkt interpretierbaren Ergebnisse zu validieren sowie mögliche Abweichungen offenzulegen. Ein solches zweischrittiges Vorgehen unter Einbezug beider Korrelationsformen wird bspw. ebenfalls von Semenova und Hassel (2015) angewendet (Semenova and Hassel 2015). Da die Spearman-Korrelation lediglich die in dieser Arbeit vorhandenen mindestens ordinal-skalierten Daten als Voraussetzung benötigt, kann dieses Maß zunächst ohne Einschränkung verwendet werden.

Als drittes Maß wird die durchschnittliche absolute Abweichung innerhalb der Kombinationen herangezogen. Das Vorgehen orientiert sich an Christensen et al. (2022) und dient insbesondere der Veranschaulichung und plakativen Darstellung der Unterschiede zwischen den Rating-Scores (Christensen et al. 2020). Um die Werte zu ermitteln, werden alle Rating-Kombinationen mit den jeweils überlappenden Scores auf Unternehmensebene separat betrachtet und in jedem Fall die absolute Abweichung pro Unternehmen ermittelt. Anschließend wird daraus pro Kombination die Summe durch die Anzahl der einbezogenen Versicherungsunternehmen geteilt, um den Durchschnitt zu bilden.

4. Analyseergebnisse

Im Rahmen der Ergebnisse beider Maße der Korrelation wurde bereits erwähnt, dass das ESG-Rating von *Franke und Bornberg* einerseits extreme Randwerte und andererseits nicht definierte Korrelationen aufweist. Grund hierfür ist mitunter, dass in einem Großteil der Kombinationen nur zwei Übereinstimmungen und damit Datenpunkte vorliegen. Mit Blick auf die Formel und aus einer statistischen Perspektive, begünstigt ein Mangel an Datenpunkten eine solche Entwicklung. Da insbesondere die Randwerte nicht repräsentativ erscheinen und um eine Ungleichbehandlung der verschiedenen Ratings zu vermeiden, folgt eine Darstellung der Mittelwerte und Mediane unter Ausschluss von *Franke und Bornberg*. Diese sind in Tabelle 3 zunächst für Bravais-Pearson ersichtlich und zeigen, dass sich Mittelwert (0,42) und Median (0,46) im Durchschnitt aller ESG-Ratings deutlich näher an die Schwelle zur mittleren Korrela-

tion bewegen. Im Rahmen der Mittelwerte fallen darüber hinaus größere Änderungen vor allem bei *Refinitiv*, *Sustainalytics*, *S&P* und *ShareAction* auf. Diese Änderungen bewirken, dass sich die Rangfolge aller Ratings gemessen am Mittelwert mit Ausnahme von *ISS* als höchstem und *Zielke* als niedrigstem Wert ändert. Zudem ändert sich die Einordnung von *Refinitiv* und *S&P* von schwach in mittel korreliert sowie in umgekehrter Weise bei *Sustainalytics*. Da der Median bereits weniger anfällig für Ausreißer wie die vorliegenden Randwerte ist, ergeben sich wahrnehmbare Unterschiede lediglich für die Ratings von *Zielke* sowie *ShareAction*. Im Falle von *Zielke* gilt nun grundsätzlich eine Einordnung als schwach negativ korreliert. Allerdings liegt auch der neue Wert mit $-0,03$ so nah an null, dass weiterhin ein nahezu unkorrelierter Zusammenhang besteht. Für *ShareAction* ist keine Änderung der Einordnung, sondern nur eine leichte Veränderung in der Rangfolge zu beobachten.

Tabelle 3

Veränderung von Mittelwert und Median der Bravais-Pearson-Korrelation ohne das Franke und Bornberg ESG-Rating

exkl. Franke u. Bornberg	#01	#02	#03	#04	#05	#06	#07	#08	#09
	Refinitiv	ISS	Sustain- alytics	Zielke	S&P	MSCI	Share Action	IVFP	Franke u. Bornberg
Mittelwert	0,57	0,58	0,46	0,01	0,53	0,33	0,44	n/a	n/a
						0,42			
Median	0,58	0,76	0,55	-0,03	0,57	0,38	0,43	n/a	n/a
						0,46			

Analog zu den Beobachtungen für die Bravais-Pearson Korrelation, sind die entsprechenden Werte für Spearman in Tabelle 4 zu sehen. Die Veränderungen im Mittelwert führen gleichermaßen zu den bereits oben beschriebenen Beobachtungen hinsichtlich der Änderung der Reihenfolge und der Einordnungen. Die einzige Ausnahme bildet das Rating von *ShareAction*, welches nun ebenfalls von schwach zu mittel korreliert wechselt. Die Werte auf Ebene des Medians zeigen, dass lediglich das Nachhaltigkeitsrating von *Refinitiv* nun einen neuen Schwellenwert überschreitet und gemäß der Definition als mittel korreliert gilt. Innerhalb der Reihenfolge der Ratings zeigen sich ebenfalls nur geringfügige Änderungen. Sowohl der jeweils übergreifende Mittelwert, als auch der Median sind identisch zu den entsprechenden Werten aus der Bravais-Pearson-Korrelation ohne *Franke und Bornberg*.

Tabelle 4

**Veränderung von Mittelwert und Median der Spearman-Korrelation
ohne das Franke und Bornberg ESG-Rating**

exkl. Franke u. Bornberg		#01	#02	#03	#04	#05	#06	#07	#08	#09
		Refinitiv	ISS	Sustain- alytics	Zielke	S&P	MSCI	Share Action	IVFP	Franke u. Bornberg
Mittelwert	0,52	0,60	0,44	0,02	0,55	0,33	0,51	n/a	n/a	
							0,42			
Median	0,53	0,79	0,53	0,01	0,56	0,34	0,46	n/a	n/a	
							0,46			

Tabelle 5

Ergebnismatrix durchschnittliche absolute Abweichung

		#01	#02	#03	#04	#05	#06	#07	#08	#09
		Refinitiv	ISS	Sustain- alytics	Zielke	S&P	MSCI	Share Action	IVFP	Franke u. Bornberg
#01	Refinitiv		34,71	15,85	22,89	15,17	12,61	54,97	19,24	11,00
#02	ISS	34,71		22,73	47,45	27,28	44,05	16,66	20,66	34,17
#03	Sustain- alytics	15,85	22,73		23,88	17,70	20,52	36,62	4,38	13,91
#04	Zielke	22,89	47,45	23,88		28,51	9,20	58,28	22,68	12,27
#05	S&P	15,17	27,28	17,70	28,51		13,45	48,83	17,57	8,50
#06	MSCI	12,61	44,05	20,52	9,20	13,45		56,08	26,13	12,86
#07	Share Ac- tion	54,97	16,66	36,62	58,28	48,83	56,08		36,91	51,60
#08	IVFP	19,24	20,66	4,38	22,68	17,57	26,13	36,91		17,84
#09	Franke u. Bornberg	11,00	34,17	13,91	12,27	8,50	12,86	51,60	17,84	
	Mittelwert	23,30	30,96	19,45	28,14	22,12	24,36	44,99	20,68	20,27
								26,03		
	Median	17,55	30,72	19,11	23,38	17,63	16,98	50,21	19,95	13,38
								23,21		

Die durchschnittlichen absoluten Abweichungen jeder Kombination sowie die Mittelwerte je Rating sind in Tabelle 5 abgebildet. Aufgrund der zuvor vorgenommenen Harmonisierung und Angleichung der Ratingskalen, können die Abweichungen auch unmittelbar inhaltlich interpretiert werden. Sämtliche Scores, welche den Abweichungen zugrunde liegen, befinden sich auf einer Skala von null bis 100. Die jeweiligen Werte veranschaulichen dementsprechend um wie viele Punkte die Scores pro Unternehmen im Schnitt auf dieser Skala voneinander abweichen. Im Gegensatz zu den Korrelationskoeffizienten liegt keine unmittelbare Zuordnung von Schwellenwerten vor. Die Darstellung erfolgt daher hauptsächlich relativ zueinander und in Teilen indikativ anhand einer Unterteilung in gleichgroße Bereiche. Auf Ebene der Mittelwerte kann zunächst ein übergreifender Wert von 26,03 festgestellt werden. Dabei ist *ShareAction* mit einem Mittelwert von 44,99 und damit der höchsten durchschnittlichen Abweichung hervorzuheben. Mit anderen Worten bedeutet dieser Wert, dass über die verschiedenen Kombinationen mit *ShareAction* hinweg ein Unternehmen durchschnittlich Scores der jeweiligen beiden Ratings erhält, welche nahezu 50 Punkte auseinanderliegen. Die nächsthöheren Werte sind im Falle von *ISS* (30,96) und *Zielke* (28,14) zu beobachten. Sämtliche weitere Ratings verteilen sich im Bereich zwischen 19,45 (*Sustainalytics*) und 24,36 (*MSCI*). Bei Betrachtung des Medians sind die drei stärksten Abweichungen weiterhin bei *ShareAction* (50,21), *ISS* (30,72) und *Zielke* (23,38) zu beobachten. Hierbei sind im Vergleich zu den Mittelwerten zwar unterschiedliche Werte zu sehen; die Reihenfolge bleibt in diesen Fällen allerdings unverändert. Die weiteren Ratings verteilen sich nun in einem gegenüber den Mittelwerten ähnlich breiten Bereich von 13,38 (*Franke und Bornberg*) bis 19,95 (*IVFP*).

Im Weiteren zeigt eine Betrachtung der einzelnen Abweichungen, dass sich circa zwei Drittel der insgesamt 36 Werte in den Bereichen zehn bis 19,99 (13 Kombinationen) und 20 bis 29,99 (9 Kombinationen) bewegen. Nahezu gleichmäßig ist die Verteilung auf die übrigen, gleichgroßen Bereiche bis zur Obergrenze von 59,99. Die vier höchsten durchschnittlichen Abweichungen mit Werten über 50 liegen entsprechend dem Mittelwert allesamt im Rahmen der Kombinationen mit *ShareAction* vor: *Zielke-ShareAction* (58,28), *MSCI-ShareAction* (56,08), *Refinitiv-ShareAction* (54,97) und *ShareAction-Franke und Bornberg* (51,60). Auf der anderen Seite des Spektrums liegen drei Kombinationen in einem niedrigen Bereich mit Werten unterhalb von zehn: *Sustainalytics-IVFP* (4,38), *S&P-Franke und Bornberg* (8,50) und *Zielke-MSCI* (9,20). Wie im Fall der Korrelationskoeffizienten lässt sich auch im Rahmen der durchschnittlichen absoluten Abweichungen beobachten, dass sich bei nahezu allen Ratings über die jeweiligen acht Kombinationen hinweg teils stark variierende Werte zeigen. Beispielsweise beinhaltet das *Zielke* ESG-Rating sowohl den höchsten als auch den dritt-niedrigsten Wert. Eine Ausnahme besteht vor allem im Rating von *ShareAction*. Hier zeichnet sich eine deutliche Tendenz zu hohen Abweichungen

ab. So besteht in der Kombination mit *ISS* eine verhältnismäßig niedrige Abweichung (16,66), wohingegen alle weiteren Werte deutlich höher zwischen 36 und 55 liegen. Dies führt zu dem hohen Mittelwert von circa 45.

Während die Ratings von *IVFP* und *Franke und Bornberg* sowie deren entsprechende Kombinationen aus den bereits beschriebenen Gründen keine oder keine zu verwendenden Korrelationen aufweisen, liegen mit den durchschnittlichen Abweichungen nun erstmals verwertbare Daten vor.

Tabelle 6
Zusammenfassende Übersicht der angewandten und relevanten Maße

	#01	#02	#03	#04	#05	#06	#07	#08	#09
	Refinitiv	ISS	Sustain- alytics	Zielke	S&P	MSCI	Share Action	IVFP	Franke u. Bornberg
Mittelwert Bravais-Pearson (exkl. Franke u. Bornberg)	0,57	0,58	0,46	v0,01	0,53	0,33	0,44	n/a	n/a
Mittelwert Spearman (exkl. Franke u. Bornberg)	0,52	0,60	0,44	0,02	0,55	0,33	0,51	n/a	n/a
Mittelwert durchschnitt- liche abs. Abweichung (inkl. Franke u. Bornberg)	23,30	30,96	19,45	28,14	22,12	24,36	44,99	20,68	20,27
Punktevergabe Korrelation	56,86	57,72	46,35	1,49	53,31	32,77	43,65	0,00	0,00
Punktevergabe durchschnittliche abs. Abweichung	76,70	69,04	80,55	71,86	77,88	75,64	55,01	79,32	79,73
Gesamtpunktzahl statistische Analyse	66,78	63,38	63,45	36,67	65,59	54,21	49,33	39,66	39,87

Tabelle 6 fasst zum Abschluss der Darstellung der ermittelten Daten die wesentlichen Daten der statistischen Maße zusammen. Aufgrund der beobachteten extremen und nicht aussagekräftigen Randwerte, werden die Korrelationsmaße in der bereinigten Form ohne Franke und Bornberg dargestellt. Um aggregierte Aussagen auf Ebene der einzelnen Ratings – und nicht auf Ebene jeder Kombination – treffen zu können, wird der jeweilige Mittelwert genutzt. An dieser Stelle werden die Werte nach Spearman zur Vollständigkeit und Übersicht aufgeführt, aber innerhalb der Diskussion überwiegend nicht weiter genutzt. Im Rahmen der durchschnittlichen absoluten Abweichungen können alle Ratings inklusive Franke und Bornberg vollständig betrachtet werden. Zwar fällt im Falle einiger Ratings auf, dass einzelne Kombinationen Ausreißer darstellen, welche sich insbesondere auf den Mittelwert des jeweiligen Ratings auswirken. Im Gegensatz zu den erwähnten Randwerten, sind diese allerdings nicht unplausibel und lassen sich nicht zwangsläufig auf mathematische Gründe zurückführen.

5. Diskussion

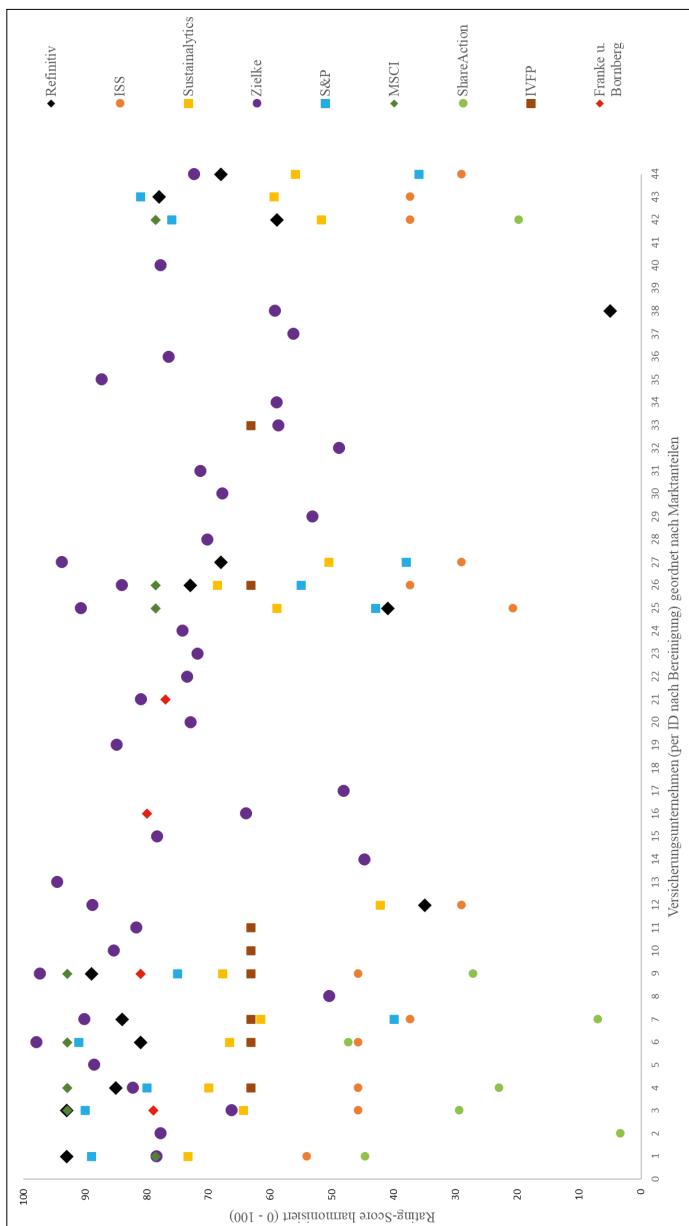


Abb. A1: Übersicht über die Verteilung der harmonisierten Rating-Scores für jedes Unternehmen

Zeitschrift für die gesamte Versicherungswissenschaft, 113 (2024) 3–4

Aus Abbildung A1 ist ersichtlich, dass die verfügbaren Scores stark voneinander abweichen. Ohne hierbei zwischen den verschiedenen Nachhaltigkeitsratings zu differenzieren, lässt sich bei nahezu allen Unternehmen beobachten, dass sich die Scores über einen großen Bereich verteilen. Das bedeutet, dass nicht nur Minimal- und Maximalwert deutlich auseinanderliegen, sondern sich auch innerhalb der Randwerte bedeutende Unterschiede ergeben. Diese Streuung birgt die Problematik, dass sich für ein einzelnes Unternehmen nicht ohne Weiteres bestimmen lässt, welches ein angemessener Score oder eine entsprechende Tendenz ist. Im Falle der *Allianz* ergeben sich bspw. Werte von 45 und 54 am unteren sowie 89 und 93 am oberen Ende mit weiteren Werten im Zwischenbereich. Unter Einbezug aller verfügbaren Scores lässt sich also eine Aussage über die ESG-Performance nicht oder nur unter Einschränkung des sehr breiten Wertebereichs treffen. Auch Aussagen über mögliche Tendenzen in den Scores über alle Versicherungsunternehmen hinweg lassen sich nicht ohne Weiteres ableiten. Zwar ist zu sehen, dass sich der überwiegende Teil aller Datenpunkte im Bereich ab 40 Punkten verteilt. Aussagen darüber hinaus sind aufgrund der sichtbaren Streuung aber nicht möglich. Ausreißer liegen bei der Betrachtung einzelner Unternehmen insbesondere am unteren Ende der Skala und lassen sich mitunter auf ratingspezifische, methodische Gründe im Rahmen des Umgangs mit fehlenden Daten zurückführen.

Im Folgenden werden die bisher getroffenen Aussagen auf Basis der Scores durch die Diskussion der Ergebnisse aus der statistischen Analyse ergänzt. Durchgeführt wurde zunächst die indikative Korrelationsanalyse mit den Koeffizienten nach Bravais-Pearson und Spearman, um die in anderen Forschungsarbeiten beobachtete fehlende Korrelation der Ratings untereinander auch im Rahmen dieser Arbeit nachzuweisen. Wichtig für die Einordnung und Diskussion ist an dieser Stelle die Referenz zu den Kreditratings, für welche eine sehr hohe Korrelation nahe eins festgestellt wird und aufgrund der grundsätzlichen Verwandtschaft zu ESG-Ratings als Vergleichswert dient (Berg, Kölbel, and Riggobon 2022). Dabei gilt allerdings, dass die Erfassung von Nachhaltigkeit mitunter stärker auf qualitativen und in Teilen subjektiveren Einschätzungen beruht. Dies bedingt, dass Unterschiede zwischen verschiedenen Ratings sowie deren Scores zwar schlüssig und prinzipiell zur Differenzierung notwendig sind, sich allerdings innerhalb eines gewissen Rahmens bewegen sollten. Ein hoher statistischer Zusammenhang sein sollte dementsprechend gegeben sein, um eine Eignung zur Erfassung der ESG-Performance festzustellen. Auf dieser Basis ergeben sich aus der Korrelationsanalyse mit Blick auf die gesamte Gruppe der Ratings einige herauszugreifende Aspekte. Wie bereits dargestellt liegt ein wesentlicher Teil der Ratingkombinationen in den Bereichen mittlerer und niedriger Korrelation, während weitere Kombinationen als unkorreliert oder schwach negativ korreliert einzurordnen sind. Dies zeigt sich auch in den einzelnen Mittelwerten der Koeffizienten auf Rating-Ebene, welche bis auf eine Ausnahme

ausschließlich im Bereich der mittleren Korrelation liegen. der Basis der Ergebnisse sowie des Gesamtdurchschnitts von 0,42 folgt die Schlussfolgerung, dass für die Gruppe der Ratings kein übergeordneter statistischer Zusammenhang festgestellt werden kann.

Wenngleich es in der Regel einen positiven Zusammenhang zwischen zwei betrachteten Ratings gibt, ist dieser häufig nicht stark ausgeprägt. Dies führt wiederum zu der Ableitung, dass zwei ESG-Ratings eine signifikant andere Sicht auf die Nachhaltigkeit desselben Unternehmens haben. Wie auch Gibson Brandon et al. (2021) feststellen und beschreiben, weisen einige in dieser Arbeit betrachtete Kombinationen dennoch hohe Korrelationswerte auf (Gibson Brandon, Krueger, and Schmidt 2021). Dies kann ein Ansatzpunkt sein, um eine tiefere Analyse und Vergleich der durch die entsprechenden Ratings angewandten Kriterien vorzunehmen. Jedoch sind auch die Werte der als stark korrelierend eingeordneten Kombinationen im Verhältnis zur nahezu perfekten positiven Korrelation der Kreditratings deutlich niedriger. Dies verdeutlicht die bereits beschriebene Problematik des nicht ausgeprägten Zusammenhangs zwischen den Ratings.

Ergänzend zur Betrachtung der Gruppe als Ganzes, lässt die Betrachtung der Mittelwerte aus den Kombinationen der Ratings Rückschlüsse auf die einzelnen Ratings zu. Im relativen Vergleich der Ratings weisen *ISS*, *Refinitiv* und *S&P* die höchsten und hier besten Werte auf. Diese sind in absoluter Betrachtung deutlich unter der Grenze zur starken Korrelation, aber liegen in der Gruppe der Ratings am oberen Ende. Im Gegensatz dazu ist das Rating von *Zielke* mit einem Wert von nahezu null und damit am unteren Ende herauszugreifen. Dieser kommt in der Berechnung insbesondere durch die teilweise vorliegende negative Korrelation zu Stande. Dies ist insbesondere deshalb zu betonen, da unter Ausschluss von *Franke und Bornberg* lediglich eine weitere Kombination verbleibt, in welcher ebenfalls ein negativer Wert vorliegt. Daraus folgt ein sehr deutlicher Unterschied von *Zielke* zu den weiteren Ratings. Ebenfalls hervorzuheben sind die Ratings von *IVFP* und *Franke und Bornberg*. Im Falle von *IVFP* können keine Korrelationskoeffizienten ermittelt werden, da alle vergebenen Scores den gleichen Wert aufweisen und die Standardabweichung null ist. Kritisch zu sehen ist dies vor dem Hintergrund, dass eine Varianz in den Scores, die ein Rating an verschiedene Unternehmen vergibt, notwendig zur Differenzierung ist. Hinzu kommt, dass auch die von *IVFP* verwendete ordinale Skala nur drei Klassen nutzt, welche eine solche mangelnde Differenzierung begünstigt. Hinzu kommt, dass die Klassenbezeichnungen mit „*Gut*“, „*Sehr gut*“ und „*Exzellent*“ allesamt positiv besetzt sind und eine tatsächliche Bewertung der ESG-Performance eines Unternehmens erschweren. *Franke und Bornberg* fallen durch Ergebnisse an beiden Rändern des Wertebereichs auf, welche sich mitunter auf die zwar gleichmäßigen aber wenigen Überschneidungen zurückführen lassen. Im Rahmen dieser Arbeit ist dies negativ zu beurteilen, da hiermit eine Beurteilung innerhalb der Gruppe nicht möglich ist.

Mit Blick auf einzelne Ratings ist des Weiteren auffällig, dass das Rating von *ShareAction* einen im Vergleich sehr hohen Mittelwert der Abweichungen aufweist. Ein möglicher Grund könnte in der hinter dem Rating stehenden Organisationsform von *ShareAction* als Stiftung (non-profit) mit Schwerpunkt auf nachhaltigkeitsbezogenen Fragestellungen und Themen liegen. Wird, bedingt durch die Ausrichtung und den Hintergrund der Organisation, ein im Gesamten anderer – eventuell strengerer – Blickwinkel auf den Komplex der Nachhaltigkeit eingenommen, kann dies zu strukturell anderen Ergebnissen führen. Trotz der im Rahmen dieser Arbeit als zunächst negativ einzustufenden Abweichungen, kann es sinnvoll sein, Ratings aus verschiedenen Organisationsformen wie beispielsweise im Falle von *ShareAction* in eine Beurteilung der ESG-Performance eines Unternehmens mit einzubeziehen. Grund ist insbesondere, dass das Aufzeigen der Abweichungen nicht inhaltlich erörtern kann, welcher Score der tatsächlich zutreffende ist. Die bereits empfohlene Bildung eines Durchschnitts könnte verschiedene Perspektiven mit mehr oder weniger strengerer Auslegung von Nachhaltigkeitskriterien zusammenführen. Schlussendlich folgt aus der Analyse der statistischen Maße, dass ein starker statistische Zusammenhang zwischen den Ratings übergeordnet nicht erkennbar ist und sich zudem deutliche durchschnittliche Abweichungen über alle Ratings hinweg ergeben. Dies bedingt, dass ein Dritter durch die hier vorliegenden ESG-Ratings nicht ohne Weiteres in die Lage versetzt werden kann, die ESG-Performance der Unternehmen in der deutschen Versicherungsbranche eindeutig zu beurteilen.

6. Fazit

Im Rahmen der vorliegenden Arbeit wurde die Forschungsfrage untersucht, ob ausgewählte Nachhaltigkeitsratings dazu geeignet sind, die ESG-Performance in der deutschen Versicherungsbranche gegenstandsangemessen zu erfassen und darzustellen. Über die Nachhaltigkeitsberichterstattung der deutschen Versicherungsunternehmen erfolgte dabei zunächst eine Zusammenstellung der Ratinglandschaft, welche entsprechend der Rahmenbedingungen dieser Arbeit auf die zu verwendenden Ratings gekürzt wurde. Für jedes verbliebene Rating wurden anschließend die Scores aller Versicherungsunternehmen – sofern vorhanden – zusammengetragen. Ferner wurden die Unterschiede in den Ratingskalen herausgearbeitet und durch eine Harmonisierung angeglichen. Infolgedessen ist eine Gegenüberstellung der Unternehmens-Scores pro Rating entstanden, in welcher die Rating-Scores unmittelbar vergleichbar sind. Auf Basis dieser Datengrundlage wurde mit statistischen Maßen indikativ untersucht, ob sich ein Zusammenhang zwischen den Ratings anhand von paarweisen Korrelationen zeigen lässt und wie stark die Rating-Scores innerhalb der Ratingpaare voneinander abweichen. Dies ist vor dem Hintergrund geschehen, dass ein fehlender Zusammenhang und große Abweichungen die Aussagekraft der Ratings

schmälern. Unsere Ergebnisse zeigen, dass die hier betrachteten ESG-Ratings gesamtheitlich betrachtet aufgrund fehlender starker statistischer Zusammenhänge sowie großer Abweichungen in den Scores und Ratingskalen nur bedingt zur Erfassung der ESG-Performance geeignet sind. Hervorzuheben sind dabei insbesondere die teils sehr hohen absoluten Abweichungen zwischen den Scores für ein einzelnes Unternehmen, was dazu führt, dass sich fundamental andere Sichtweisen auf die Nachhaltigkeit des Unternehmens ableiten lassen.

Die vorliegenden Ergebnisse ergänzen den bestehenden Forschungsstand um die spezifische Betrachtung der Versicherungsbranche im deutschen Markt anhand aktueller Daten sowie neu aufgestellter und durch die EU-Studie begründeter Kriterien. Mit Blick auf das Vorgehen und die Interpretation der Ergebnisse ist es notwendig, auf einige Einschränkungen hinzuweisen. Im Kontext der Rating-Scores fällt auf, dass es Unterschiede auf den betrachteten Unternehmensebenen gibt. Während einige Ratings auf das Versicherungsunternehmen als Ganzes schauen, fokussieren andere Ratings bspw. explizit die entsprechende Lebensversicherungsgesellschaft. Hierdurch kann potentiell eine geringe Unschärfe entstehen. Des Weiteren kann es zwischen den Ratings durch unterschiedliche Aktualisierungszyklen zu zeitlich nicht synchronisierten Beurteilungen der Unternehmen kommen. Durch das Inkrafttreten der Corporate Sustainability Reporting Directive ab der Berichterstattung über das Jahr 2023 müssen Unternehmen künftig umfassender und nach einheitlicheren Maßstäben verpflichtend im Lagebericht über ihre Nachhaltigkeit berichten. Durch eine stärkere Quantifizierung der Berichtsinhalte sowie eine externe Prüfung der Nachhaltigkeitsberichterstattung wird eine erhöhte Verlässlichkeit für die Nutzer der Nachhaltigkeitsinformationen angestrebt. Eine erneute bzw. weiterführende Untersuchung der Ratings, welche die in der nächsten Zeit veröffentlichten Daten nutzen, wird Aufschluss über den Informationsgehalt der Nachhaltigkeitsberichterstattung und den daraus erstellten Ratings liefern.

Anhang
Gesamtübersicht final einbezogener Versicherungsunternehmen

Tabelle A.1

**Auflistung aller nach der Bereinigung verbleibender Unternehmen
unter Angabe der jeweiligen Marktwerte**

ID nach Bereinigung	Name Versicherungsunternehmen	Marktanteil in %
1	Allianz	17,81
2	R+V	6,84
3	Generah	6,25
4	ERGO	5,68
5	Debeka	5,42
6	AXA	4,87
7	Talanx	4,06
8	HUK Coburg	3,57
9	Zurich	2,57
10	SIGNALIDUNA	2,54
11	ALTE LEIPZIGER- HALLESCHE	2,03
12	W&W	2,02
13	Gothaer	1,93
14	Cantirrentale	1,89
15	LVM	1,75
16	Nürnberger	1,50
17	VHV	1,46
18	Viriclium	1,44
19	HanseMerkur	1,40
20	DEVK	1,32
21	Bannen:ia	1,13
22	VOLKSWOHL BUND	0,74
23	ARAG	0,66
24	WWK	0,60
25	Basler	0,56
26	Sclr.veizerische Rente:namtalt/Swiss Life	0,55
27	Helvetia	0,41
28	INTER	0,41
29	SDK	0,41
30	Con::ord:ia	0,38
31	Universa	0,37
32	WGV	0,36
33	Stuttgarter	0,35

ID nach Bereinigung	Name Versicherungsunternehmen	Marktanteil in %
34	Münchener Verein	0,35
35	Die Bayerische	0,31
36	Itzehoer	0,28
37	Mecklenburgische	0,28
38	RheinLarrl	0,27
39	IDEAL	0,23
40	VPV	0,20
41	Wertgarantie	0,17
42	MSIG	0,16
43	Verti	0,16
44	InterRisk	0,11
Summe repräsentierter Marktanteil		85,80

Quelle: Kölner Institut für Versicherungsinformation und Wirtschaftsdienste 2023.

Literaturverzeichnis

- Berg, F./Kölbl, J. F./Rigobon, R. (2022): Aggregate Confusion: The Divergence of ESG Ratings*. In: *Review of Finance* 26(6), 1315–44. DOI: <https://doi.org/10.1093/rof/rfac033>.
- Chatterji, A. K./Durand, R./Levine, D. I./Touboul, S. (2016): Do Ratings of Firms Converge? Implications for Managers, Investors and Strategy Researchers. In: *Strategic Management Journal* 37, 1597–1614. DOI: <https://doi.org/10.1002/smj>.
- Christensen, D./Serafeim, G./Sikochi, A./Bradshaw, M./Carter, M. E./Doyle, J./Gillette, J. et al. (2020): Why Is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings We Thank Vishal Baloria. In: Social Science Research Network 617.
- Europäisches Parlament (2022): Richtlinie (EU) 2022/2464. Amtsblatt Der Europäischen Union 1999 (401), 15–80.
- Gibson Brandon, R./Krueger, P./Schmidt, P. S. (2021): ESG Rating Disagreement and Stock Returns. In: *Financial Analysts Journal* 77(4), 104–27. DOI: <https://doi.org/10.1080/0015198X.2021.1963186>.
- Kölner Institut für Versicherungsinformation und Wirtschaftsdienste (2023): Die Marktanteile der Anbieter auf dem deutschen Erstversicherungsmarkt im Jahr 2022 und im Zeitvergleich der Jahre 2015 bis 2022 – Marktanteilsstatistik –.
- Semenova, N./Hassel, L. G. (2015): On the Validity of Environmental Performance Metrics. In: *Journal of Business Ethics* 132(2), 249–58. DOI: <https://doi.org/10.1007/s10551-014-2323-4>.
- Serafeim, G./Yoon, A. (2021): Which Corporate ESG News Does the Market React To?. In: Harvard Business School Working Paper 21–115. DOI: <https://doi.org/10.2139/ssrn.3832698>.