



Kalibrierung Logib Modul 2 durch Umlegung von Modul 2 Kontrolldaten ins Modul 1 Bericht

BEREICH ARBEIT

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra



Eidgenössisches Departement des Innern EDI
Eidgenössisches Büro für die Gleichstellung von Frau und Mann EBG

IMPRESSUM

TITEL

Kalibrierung Logib Modul 2 durch Umlegung von
Modul 2 Kontrolldaten ins Modul 1

AUTOR

Dr. Florian Chávez-Juárez, CORESO - Collabroative Research
Solutions Sàrl

HERAUSGEBERIN

Eidgenössisches Büro für die Gleichstellung
von Frau und Mann EBG

VERTRIEB

Eidgenössisches Büro für die Gleichstellung
von Frau und Mann EBG
Schwarztorstrasse 51, 3003 Bern
ebg@ebg.admin.ch
www.ebg.admin.ch

Dieser Bericht wurde im Auftrag des EBG verfasst. Die darin ent-
haltenen Einschätzungen und Interpretationen entsprechen nicht
zwingend der Sicht des EBG.



Kalibrierung Logib Modul 2 durch Umlegung von Modul 2 Kontrolldaten ins Modul 1

Dr. Florian Chávez Juárez
Fassung vom 13. Mai 2024

Impressum

Titel:

Kalibrierung Logib Modul 2 durch Umlegung von Modul 2 Kontrolldaten ins Modul 1

Auftraggeberin:

Eidgenössisches Büro für die Gleichstellung von Frau und Mann (EBG)

Autor:

Dr. Florian Chávez-Juárez, CORESO - Collaborative Research Solutions Sàrl.
florian.chavez@coreso.ch

Die in diesem Dokument enthaltenen Ergebnisse, Interpretationen und Schlussfolgerungen geben nicht notwendigerweise die Meinung der Auftraggeberin wieder.

Inhaltsverzeichnis

1	Einleitung	5
2	Mapping der Funktionsbewertung M2 in ANF/BST M1	6
2.1	Inhaltliches Mapping	6
2.2	Mathematisches Mapping	7
2.3	Gleichung mit Funktionswert statt ANF, BST	8
2.4	Zusammenfassung der Umlegungsvarianten	8
3	Daten	9
4	Resultate	10
4.1	Vergleich der Umlegungen	10
4.2	Resultate zu den spezifischen Fragen des Mandats	11
4.2.1	Vergleich der Resultate M1 vs. M2: Scatterplot	12
4.2.2	Unterschiedliche Vorzeichen M1 vs. M2	14
4.2.3	Verteilung der Resultate nach Geschlecht	17
4.2.4	Wo muss der Grenzwert festgelegt werden?	18
5	Leave-one-out Ansatz	20
5.1	Idee und Umsetzung	20
5.2	Deskriptive Statistik	21
5.3	Resultate zu den spezifischen Fragen mit dem Leave-one-out Ansatz	24
5.3.1	Vergleich der Resultate M1 vs. M2: Scatterplot	24
5.3.2	Unterschiedliche Vorzeichen M1 vs. M2	26
5.3.3	Verteilung der Resultate nach Geschlecht	28
5.3.4	Wo muss der Grenzwert festgelegt werden?	28
6	Schlussbemerkungen	31
A	Resultate anderer Umlegungen	34
B	Resultate anderer Umlegungen (Leave-one-out Ansatz)	40

Abbildungsverzeichnis

1	Kontrollen nach Anzahl Mitarbeitenden und Anzahl Frauen	9
2	Korrelation zwischen den verschiedenen Umlegungen	10
3	Verteilung der R2-Statistik nach Umlegung	11
4	Korrelation zwischen dem Gesamtscore M2 und dem Geschlechterkoeffizienten M1	11
5	Vergleich der Resultate M1 und M2	13
6	Verteilung der Resultate nach Geschlecht	17
7	Kumulative Verteilung der M2 Kontrollen	19
8	Anteil der Kontrollen mit gleichem Resultat M1 und M2	20
9	Vergleich des normalen Scores mit dem LOO-Scores	21
10	Verteilung des LOO-Gewinns	22
11	Minimale und maximale LOO-Scores	23
12	Vergleich der Resultate M1 und M2 (LOO-Ansatz)	25
13	Verteilung der Resultate nach Geschlecht	28
14	Kumulative Verteilung der M2 Kontrollen	29
15	Anteil der Kontrollen mit gleichem Resultat M1 und M2	30
16	Hauptresultat am Beispiel der Umlegung FW1	31
17	Hauptresultat am Beispiel der Umlegung FW1 (Leave-one-out)	32
18	Vergleich der Resultate M1 und M2	34
19	Verteilung der Resultate nach Geschlecht	37
20	Kumulative Verteilung der M2 Kontrollen	38
21	Anteil der Kontrollen mit gleichem Resultat M1 und M2	39
22	Vergleich der Resultate M1 und M2 (LOO-Ansatz)	40
23	Kumulative Verteilung der M2 Kontrollen (LOO-Ansatz)	42
24	Anteil der Kontrollen mit gleichem Resultat M1 und M2 (LOO-Ansatz)	43

Tabellenverzeichnis

1	Vorschlag gemäss Gajta (2023)	6
2	Vorschlag Mapping Methode 3a	7
3	Zusammenfassung der Varianten	8
4	Anzahl berücksichtigter Kontrollen nach Quelle und Jahr	9
5	Anteil Kontrollen mit Vorzeichenwechsel	14
6	Anzahl Kontrollen mit Vorzeichenwechsel (detailliert)	16
7	Deskriptive Statistik zum Leave-one-out-Ansatz	22
8	Anzahl Kontrollen mit Vorzeichenwechsel (detailliert), LOO	27
9	Anteil Kontrollen mit Vorzeichenwechsel	35
10	Anzahl Kontrollen mit Vorzeichenwechsel (detailliert)	36
11	Anzahl Kontrollen mit Vorzeichenwechsel (detailliert, LOO-Ansatz))	41

1 Einleitung

Im Juni 2023 hat das *Eidgenössisches Büro für die Gleichstellung von Frau und Mann (EBG)* Anpassungen an der Toleranzschwelle in Logib Modul 1 per Januar 2024 beschlossen ([EBG, 2023a](#)). Diese Anpassungen, insbesondere die Abschaffung eines zweiten statistischen Tests, führen zu einer Veränderung der Kriterien zur Einhaltung der Lohnungleichheit. Aus diesem Grund sollen nun auch die Kriterien im Modul 2 von Logib überprüft und gegebenenfalls angepasst werden, damit eine Gleichbehandlung der Unternehmungen in beiden Module gewährleistet werden kann.

Das EBG hat Fachpersonen zu möglichen methodischen Ansätzen für diese Überprüfung der Kalibrierung des Moduls 2 befragt und dabei kam die Idee einer Umlegung von Lohndaten aus Modul 2 ins Modul 1 auf. Die Idee besteht darin, plausibilisierte Lohndaten aus Modul-2-Kontrollen so aufzubereiten, dass sie auch für eine Analyse mit Modul 1 geeignet sind. Dadurch können die Ergebnisse beider Module für die gleichen Unternehmen verglichen und Modul 2 so kalibriert werden, dass die Ergebnisse möglichst übereinstimmen.

In diesem Dokument wird die Umsetzung dieser Idee beschrieben. Dabei ist zu beachten, dass die vorliegende Version des Dokuments das Ergebnis eines iterativen Prozesses zwischen dem Autor, dem EBG und weiteren Fachpersonen im Bereich Lohnungleichheit ist, was sich auch in den einzelnen Abschnitten des Dokuments widerspiegelt. In einer ersten Iteration wurden die inhaltlichen und mathematischen Grundlagen der Umlegung erarbeitet und mit dem EBG sowie weiteren Fachpersonen diskutiert (siehe Abschnitt 2).

Anschliessend wurden die Umlegungen für die verfügbaren Daten (siehe Abschnitt 3) vorbereitet und durchgeführt. Die Ergebnisse (siehe Abschnitt 4) wurden wiederum mit dem EBG und weiteren Fachpersonen diskutiert. Aus diesen Diskussionen resultierte der Vorschlag, die Volatilität der Modul-2-Analysen durch den Leave-one-out-Ansatz zu reduzieren. Dieser Vorschlag wurde in der Folge umgesetzt und ist hier im Abschnitt 5 dargestellt.

Die Schlussbemerkungen im Abschnitt 6 beziehen sich dann auf den ganzen Prozess und alle Resultate.

2 Mapping der Funktionsbewertung M2 in ANF/BST M1

Um eine Lohnvergleichsanalyse mit Logib durchzuführen, müssen die Lohndaten gemäss der Wegleitung¹ des jeweiligen Moduls aufbereitet werden. Sämtliche individuellen Angaben zu den Personen sowie deren Vergütungen sind in beiden Modulen identisch. Der einzige Unterschied in der Aufbereitung der Daten zwischen Modul 1 und Modul 2 besteht in der Bewertung der Funktionen. In Modul 1 wird jede Funktion in zwei Dimensionen bewertet: berufliche Stellung (BST) und Anforderungsniveau (ANF). Im Gegensatz dazu müssen die Funktionen in Modul 2 in 6 verschiedenen Dimensionen bewertet werden.

Um plausibilisierte Lohndaten aus Modul 2 in einer Modul-1-Analyse verwenden zu können, müssen die Bewertungen in den 6 Dimensionen von Modul 2 auf die beiden Dimensionen des Moduls übertragen werden. Dazu sind sowohl inhaltliche als auch mathematische Überlegungen anzustellen.

2.1 Inhaltliches Mapping

Gajta (2023) präsentiert einen Entwurf für das Mapping der Funktionsbewertungen des Modul 2 in die beiden Variablen ANF (Anforderungsniveau) und BST (Berufliche Stellung) des Modul 1. Gajta (2023) sieht folgenden Vorschlag für das Mapping vor:

Tabelle 1: Vorschlag gemäss Gajta (2023)

Funktionsbewertung Modul 2	Mapping in Variable M1
Anforderung an die Ausbildung	ANF oder BST oder weglassen, da diese Anforderung über die Variable "Ausbildung" im Modul abgedeckt wird.
Anforderungen an die Autonomie	ANF
Anforderungen an spezifisches Fach- und Methodenwissen	ANF
Psychische und soziale Anforderungen und Belastungen	ANF
Verantwortungsbezogene Anforderungen und Belastungen	BST
Körperliche Anforderungen und Belastungen	Weglassen, da die körperliche Belastung im Modul 1 nicht berücksichtigt wird.

Gemäss diesem Vorschlag ist einzig die Behandlung der "Anforderungen an die Ausbildung" noch nicht klar. Unserer Einschätzung nach ist es wenig zielführend, diese Dimension in der Variable BST zu berücksichtigen. Die Variable BST entspricht gemäss der Wegleitung (EBG, 2023b) ziemlich genau der Dimension "Verantwortungsbezogene Anforderungen und Belastungen" in Modul 2 (EBG, 2023c). In beiden Wegleitungen werden Führungsverantwortung², Fachverantwortung, Verantwortung für menschliches Leben und für wertvolle Güter und Materialien erwähnt. Zudem werden beide Variablen auf einer Skala mit 5 möglichen Werten bewertet. Somit wird hier ein 1:1 Mapping der Variable "Verantwortungsbezogene Anforderungen und Belastungen" auf BST genutzt und keine weiteren Faktoren zu berücksichtigt. Die Ausbildung wird in der Wegleitung Modul 1 (EBG, 2023b) bei der Variable BST nicht erwähnt.

Somit stellt sich noch die Frage, ob die Dimension "Anforderungen an die Ausbildung" in die Variable ANF aufgenommen werden soll oder ganz weggelassen werden kann. Die Wegleitung zu Modul 1 besagt, dass die "typischerweise geforderte Ausbildung" Teil der Variable ANF ist, was dafür spricht, sie einzubeziehen. Diese Ansicht wurde durch die Konsultation weiterer Fachpersonen bestätigt. Es stellte sich heraus, dass es aus arbeitswissenschaftlicher Sicht notwendig ist, die Variable ANF einzubeziehen, da sie das intellektuelle Niveau einer Funktion erfasst. Die eigentliche Ausbildung der Funktionstragenden, wie sie in Modul 1 behandelt wird, deckt dieses Niveau nicht ab.

Im Rahmen der Konsultation wurde auch die Frage aufgeworfen, ob die Dimension "Autonomie" gegebenenfalls in die Variable BST aufgenommen werden könnte, da die Anzahl und der Umfang der selbständig zu treffenden

¹Siehe EBG (2023b) für Modul 1 und EBG (2023c) für Modul 2

²Genauer Wortlaut ist *Führungshierarchie im Unternehmen* (Modul 1) und *Führungsverantwortung* (Modul 2)

Entscheidungen in engem Zusammenhang mit der beruflichen Stellung (BST) stehen. Ob diese Dimension besser in der Variable ANF oder BST berücksichtigt wird, kann von Betrieb zu Betrieb unterschiedlich sein und müsste im Einzelfall beurteilt werden. Eine solche Einzelfallbeurteilung ist im Rahmen dieser Studie nicht möglich, jedoch können beide Optionen als Varianten in die Berechnungen einbezogen werden.

Somit werden folgende inhaltlichen Umlegungen vorgeschlagen:

Tabelle 2: Vorschlag Mapping Methode 3a

Funktionsbewertung Modul 2	Umlegung 1	Umlegung 2	Umlegung 3
Anforderung an die Ausbildung	X	ANF	ANF
Anforderungen an spezifisches Fach- und Methodenwissen	ANF	ANF	ANF
Psychische und soziale Anforderungen und Belastungen	ANF	ANF	ANF
Anforderungen an die Autonomie	ANF	ANF	BST
Verantwortungsbezogene Anforderungen und Belastungen	BST	BST	BST
Körperliche Anforderungen und Belastungen	X	X	X

Bemerkungen: X bedeutet, dass dieser Faktor nicht berücksichtigt wird.

2.2 Mathematisches Mapping

Bei den Umlegungen 1 und 2 ist die Berechnung bei der *beruflichen Stellung* (BST) trivial, da beide Variablen praktisch identisch definiert werden. Um ganz korrekt zu sein, muss die Variable jedoch umgedreht werden³:

$$BST = 6 - F_{\text{Verantwortung}} \quad (1)$$

Bei der Variable ANF, sowie bei BST in Umlage 3, ist die Ausgangslage etwas komplizierter, da mehrere kategoriale Variablen kombiniert und dann wieder auf eine kategoriale Variable umgelegt werden müssen. [Gajta \(2023\)](#) schlägt einen Ansatz mit Mittelwerten vor. Wir sind auch der Meinung, dass ein solcher Ansatz zielführend ist, möchten aber zwei mathematische Optionen vorschlagen:

1. Einen einfachen, normalisierten⁴ Mittelwert und dann eine Rundung zur nächsten ganzen Zahl.
2. Einen Mittelwert, der so normalisiert wird, dass sicher alle möglichen Werte der Variable ANF benutzt werden.

Mathematisch können die beiden Vorschläge wie folgt umgesetzt werden. In einem ersten Schritt wird ein normalisierter Mittelwert berechnet:

$$\bar{f}_n = \frac{\bar{f} - 1}{5 - 1} \quad (2)$$

$$\tilde{f}_n = \frac{\bar{f} - \min(\bar{f})}{\max(\bar{f}) - \min(\bar{f})} \quad (3)$$

wobei $\bar{f}$ der einfache Mittelwert der berücksichtigten Faktoren ist. $\bar{f}_n$ ist dann der normalisierte Mittelwert auf einer Skala von 0 bis 1. Bei dieser Berechnung kann es vorkommen, dass sich alle Werte irgendwo auf dieser Skala konzentrieren. $\tilde{f}_n$ ist ebenfalls ein normalisierter Mittelwert, jedoch wird mit dem effektiven Minimum und Maximum normalisiert, was dazu führt, dass sicher die ganze Skala von 0 bis 1 benutzt wird.

³Dies ist für die Berechnung von Logit eigentlich irrelevant, da die Variable sowieso als 4 Dummy Variablen in die Regression genommen wird.

⁴Unter 'normalisiert' verstehen wir die Normalisierung auf den möglichen Wertebereich zwischen 1 und 4 der Variable ANF

In einem letzten Schritt werden dann diese beiden Variablen zurück in die kategoriale Variable ANF gewandelt:

$$ANF = \begin{cases} 1 & \text{wenn } 1.00 \geq \bar{f}_n > 0.75 \\ 2 & \text{wenn } 0.75 \geq \bar{f}_n > 0.50 \\ 3 & \text{wenn } 0.50 \geq \bar{f}_n > 0.25 \\ 4 & \text{wenn } 0.25 \geq \bar{f}_n \geq 0.00 \end{cases} \quad (4)$$

Für $\bar{f}_n$ wird genau die gleiche Umwandlung benutzt. Bei der Variable BST in der Umlegung 3 wird analog vorgegangen, einfach über 5 Stufen.

Verworfenen Idee: Maximalwert

Neben dem Mittelwert wurde auch überlegt, den Maximalwert der berücksichtigten Dimensionen zu verwenden. Dies hätte den Vorteil, dass es sich weiterhin um eine kategoriale Variable handeln würde. Allerdings würde die Variable ANF dann jeweils nur durch eine der Dimensionen definiert und innerhalb des Unternehmens nicht unbedingt immer durch dieselbe Dimension. Die Berücksichtigung aller Dimensionen erscheint logischer und näher an den Erläuterungen zu ANF in der Wegleitung (EBG, 2023b).

2.3 Gleichung mit Funktionswert statt ANF, BST

Neben dem Mapping, wie zuvor beschrieben, soll auch ein etwas modifiziertes Logib Modul 1 berechnet werden, wobei die beiden Variablen BST und ANF durch den Funktionswert aus Modul 2 ersetzt werden. Dies hat den grossen Vorteil, dass wir uns auf die in Modul 2 bestehende Kombination der verschiedenen Faktoren berufen können. Diese sind für jede Unternehmung optimiert und sollten demnach besser geeignet sein, die Funktion zu bewerten. Wie in Gajta (2023) vorgeschlagen, soll der Funktionswert linear und im Quadrat in die Regression einfließen. Zusätzlich wird auch eine Version mit Dummy Variablen für jeden Funktionswert geschätzt, sofern genügend Freiheitsgrade zur Verfügung stehen.

2.4 Zusammenfassung der Umlegungsvarianten

Die Tabelle 3 fasst die insgesamt 8 vorgeschlagenen Umsetzungsvarianten zusammen.

Tabelle 3: Zusammenfassung der Varianten

Variante	Umlegung	Normalisierung	Logib
UL1a	Umlegung 1	Mit theoretischem min/max	Standard
UL1b	Umlegung 1	Mit effektivem min/max	Standard
UL2a	Umlegung 2	Mit theoretischem min/max	Standard
UL2b	Umlegung 2	Mit effektivem min/max	Standard
UL3a	Umlegung 3	Mit theoretischem min/max	Standard
UL3b	Umlegung 3	Mit effektivem min/max	Standard
FW1	ersetzt durch FW	–	Mit FW linear und im Quadrat
FW2	ersetzt durch FW	–	Mit FW als Dummy

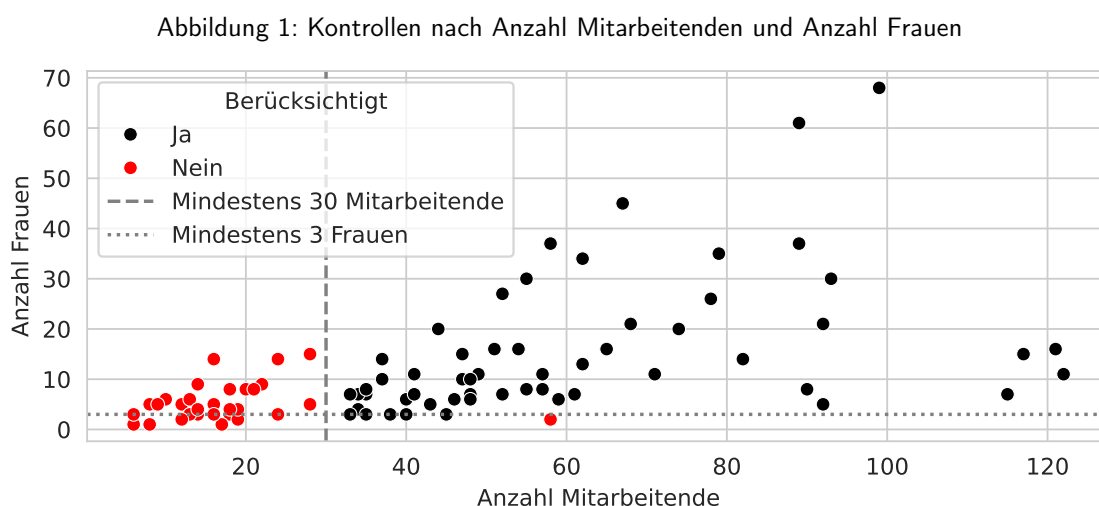
FW = Funktionswert (aus Modul 2)

Dabei handelt es sich um 3 verschiedene Umlegungen mit je 2 Normierungsoptionen. Zusätzlich werden noch zwei Umsetzungsvarianten auf der Basis des Funktionswertes von Modul 2 berücksichtigt, die sich nur in der funktionalen Form der Regression unterscheiden.

3 Daten

Die aktuellen Auswertungen basieren auf insgesamt 88 Modul-2-Kontrollen, welche in anonymisierter Form vom EBG zur Verfügung gestellt wurden. Sie stammen aus Kontrollen des EBG und anderer Kontrollbehörden, welche die Daten für Forschungszwecke dem EBG zugänglich gemacht haben. Alle Kontrollen wurden als Export-Files zur Verfügung gestellt und erneut in das aktuelle Logib-Online-Tool eingelesen, um sicherzustellen, dass alle Auswertungen auf der gleichen Berechnungsgrundlage basieren.

Die Abbildung 1 zeigt die Verteilung der Unternehmensgrößen in den Daten. Auf der horizontalen Achse wird die Anzahl Mitarbeitende und auf der vertikalen Achse die Anzahl Frauen ausgegeben.



Um den Bedenken gegenüber Regressionen mit sehr wenigen Datenpunkten Rechnung zu tragen und gleichzeitig die Anzahl Resultate in diesem Dokument auf einem sinnvollen Niveau zu halten, wurden nur Unternehmen mit mindestens 30 Mitarbeitenden und mindestens 3 Mitarbeitenden pro Geschlecht berücksichtigt. In der Abbildung 1 sind die in der Berechnung berücksichtigten Unternehmen schwarz und die nicht berücksichtigten rot markiert.

Die Tabelle 4 zeigt die Anzahl der in der Analyse berücksichtigten Kontrollen nach Jahr und Quelle auf.

Tabelle 4: Anzahl berücksichtigter Kontrollen nach Quelle und Jahr

Jahr	2018	2019	2020	2021	2022	2023	Total
EBG-Kontrollen	0	0	0	20	18	0	38
Kontrollen anderer Kontrollbehörden	3	0	1	4	9	2	19
						Total	57

Für gewisse Resultate (z.B. Abbildung 5) wurden statistische Signifikanztests (immer mit $H_0 : \beta = 0$) durchgeführt. Es ist zu beachten, dass die statistische Aussagekraft dieser Tests aufgrund des geringen Stichprobenumfangs sehr begrenzt ist. Ein nicht signifikanter Effekt könnte signifikant werden, wenn die Stichprobe ein wenig grösser wäre.

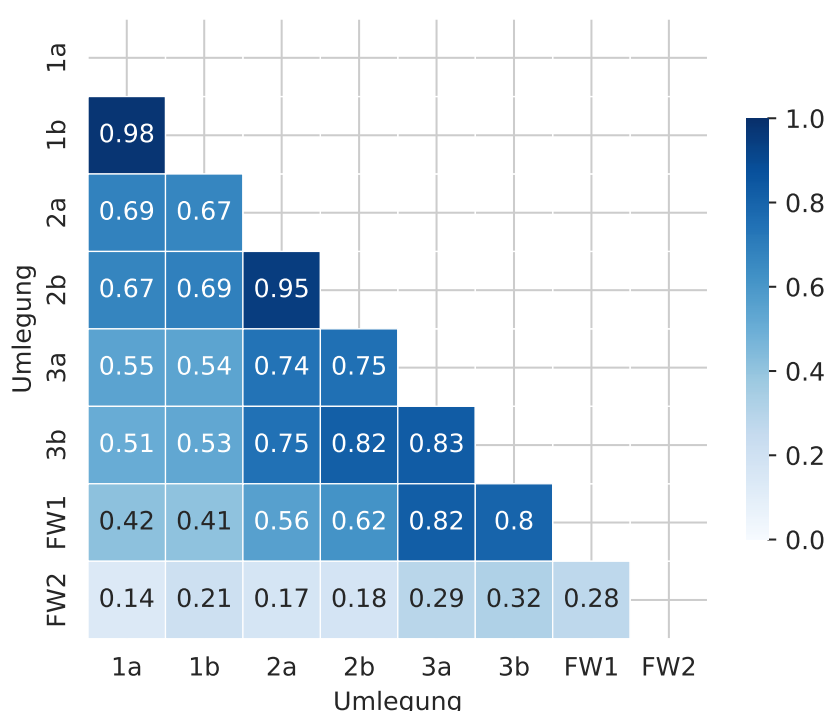
4 Resultate

In diesem Abschnitt gehen wir zuerst auf die verschiedenen Umlegungen ein. Dabei wird auch eine Auswahl vorgenommen und die Hauptresultate nur für diese ausgewählten Umlegungen präsentiert. Die restlichen Resultate befinden sich im Anhang A. Im zweiten Teil des Abschnitts werden dann die 4 Fragen zu diesem Mandat einzeln behandelt.

4.1 Vergleich der Umlegungen

In einem ersten Schritt sollte angeschaut werden, ob die verschiedenen Umlegungen überhaupt zu unterschiedlichen Resultaten in Bezug auf den Geschlechterkoeffizienten β_{gender} führen. Die Abbildung 2 zeigt die lineare Korrelation (Pearson) zwischen den geschätzten Werten von β_{gender} nach Umlegungsart (siehe Tabelle 3 für die Definition).

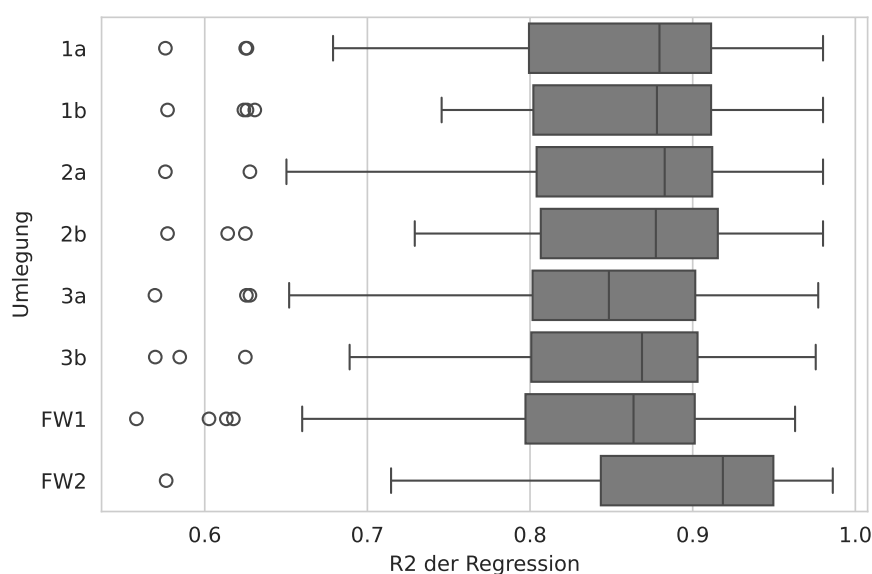
Abbildung 2: Korrelation zwischen den verschiedenen Umlegungen



Dabei fällt auf, dass die mathematische Umlegung (a vs. b) bei den Umlegungen 1 und 2 wenig ausmacht, während sie bei der Umlegung 3 mehr zum Tragen kommt. Die Umlegungen 2 und 3 sind stark korreliert, wobei die Umlegung 3 stärker mit der Schätzung FW1 (Funktionswert als Polynom) korreliert als die Umlegung 2. Hingegen korreliert das Resultat der Schätzung FW2 (Funktionswert als Dummy) nur wenig mit den anderen Varianten.

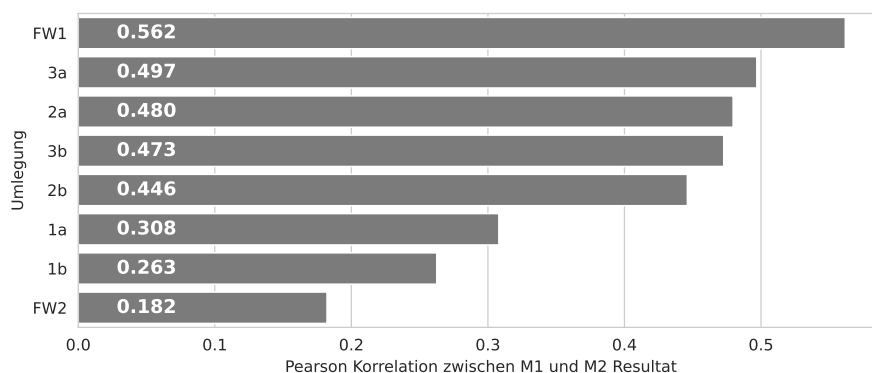
Die Schätzung FW2 verhält sich auch sonst ziemlich unterschiedlich, wie man in der Abbildung 3 klar sehen kann. Die Abbildung zeigt die Verteilung der R²-Statistik der geschätzten Regressionsmodelle. Während alle anderen Umlegungen ähnliche Verteilungen aufweisen, beobachten wir bei FW2 markant höhere R²-Statistiken. Dies könnte auf den ersten Blick für die Umlegung FW2 sprechen.

Abbildung 3: Verteilung der R2-Statistik nach Umlegung



Betrachtet man jedoch die Korrelation der Ergebnisse zwischen M1 und M2 (Abbildung 4), so wird deutlich, dass die Umlegung FW2 wesentlich schlechter abschneidet als die anderen.

Abbildung 4: Korrelation zwischen dem Gesamtscore M2 und dem Geschlechterkoeffizienten M1



Unserer Einschätzung nach deutet dies auf einen Overfit durch die Umlegung FW2 hin, da sehr viele Dummy-Variablen ins Modell miteinbezogen werden. Wenn ein Geschlecht auf einige wenige Werte konzentriert ist, kann dies einen Teil der geschlechtsspezifischen Effekte aus dem Modell herausnehmen. Aus diesem Grund empfehlen wir, die Umlegung FW2 nicht weiter zu berücksichtigen.

Weiter können wir auch sehen, dass die Umlegung 1 (a und b) merklich tiefere Korrelationen erzeugt, was sich mit der Einschätzung anderer Fachpersonen im Rahmen einer Konsultation deckt. Somit empfehlen wir, auch diese Umlegung nicht weiter zu berücksichtigen. Um die Anzahl der Varianten noch etwas weiter zu reduzieren, werden wir auch die Resultate der Umlegung 2b nicht ausführlich diskutieren. Sämtliche Resultate sind jedoch im Anhang A verfügbar.

4.2 Resultate zu den spezifischen Fragen des Mandats

Wir gehen nun auf die 4 spezifischen Fragen des Mandats ein. Dabei haben wir die Reihenfolge der Fragen etwas umgedreht, um zuerst einen Einblick in die Schätzungen zu geben und dann auf die Hauptfrage der

Kalibrierung zu kommen.

4.2.1 Vergleich der Resultate M1 vs. M2: Scatterplot

Fragestellung:

Gibt es einen Zusammenhang zwischen dem Geschlechterkoeffizienten M1 und dem Risikoscore M2?

Um einen Scatterplot mit dem Risikoscore auf der vertikalen und dem Geschlechterkoeffizienten auf der horizontalen Achse erstellen zu können, musste das Gesamtscore Modul 2 leicht angepasst werden. Dies ist nötig, weil der Score M2 immer positiv ist und mit einem Zusatz (zu Ungunsten von...) versehen ist, während der Score Modul 1 das Vorzeichen (negativ = zu Ungunsten von Frauen) nutzt. Aus diesem Grund wurde der Score wie folgt angepasst:

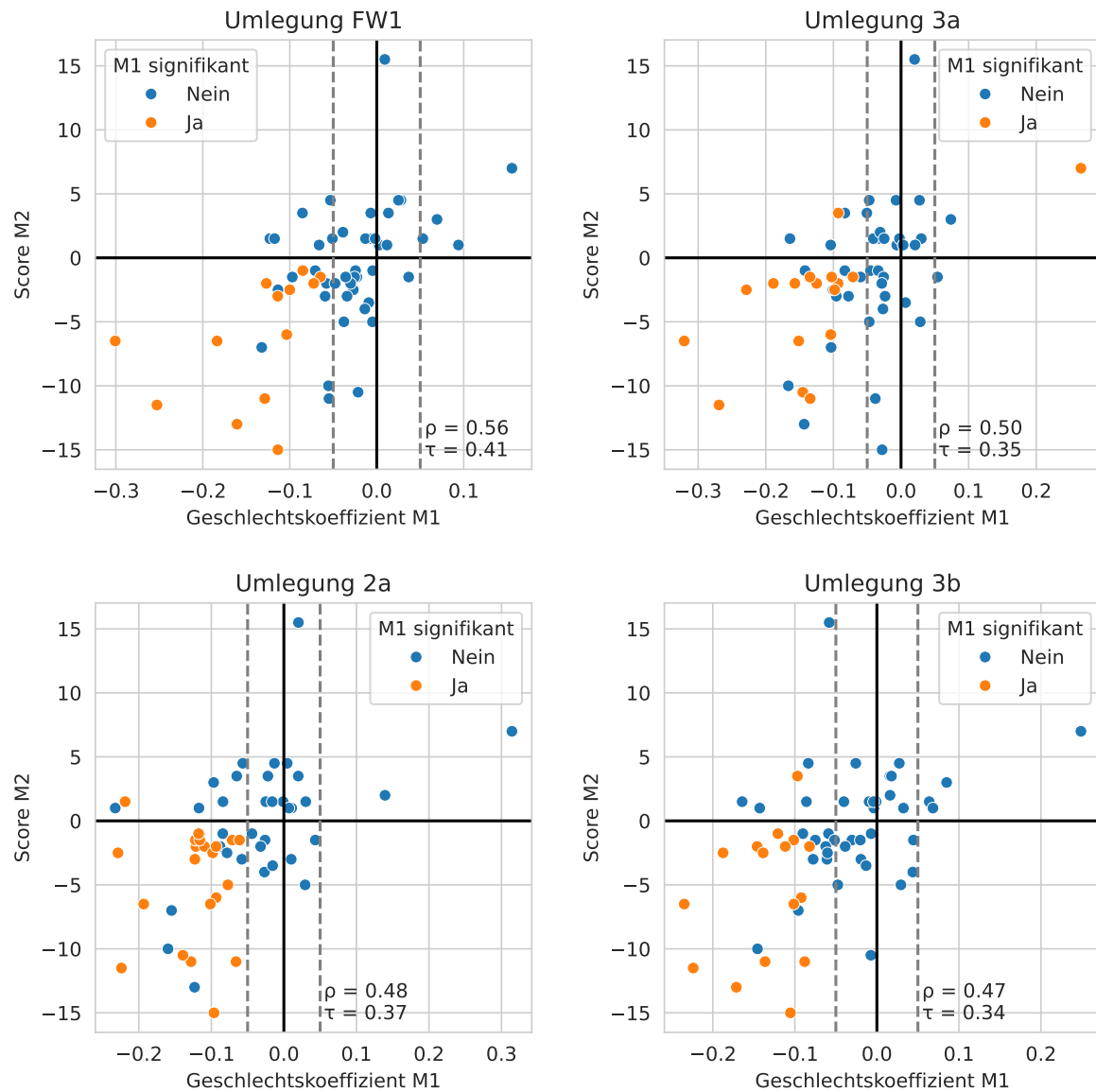
1. Bei tiefen Scores wird nicht angezeigt, ob es zu Ungunsten von Frauen oder Männern ist. Dies führt aber dazu, dass im Scatterplot alle bei 1 (im positiven Bereich) liegen würden. Deshalb wurde bei allen Kontrollen der Score noch einmal berechnet und auch bei diesen Scores eruiert, ob es zu Ungunsten der Frauen oder Männer ausfällt. Wenn die beiden Werte identisch waren, wurde ein Wert von 0 definiert.
2. Alle Scores zu Ungunsten der Frauen wurden mit -1 multipliziert.

Nach diesen Anpassungen ergaben sich die Scatterplots, die in Abbildung 5 gezeigt werden. Die horizontale Achse zeigt den geschätzten Geschlechterkoeffizienten Modul 1 und die vertikale Achse den Score aus Modul 2. Die Farben beziehen sich auf die Signifikanz des Geschlechterkoeffizienten. Dabei ist jedoch Vorsicht geboten, da es zum Teil sehr kleine Unternehmungen sind und somit eine Signifikanz bei 95%-Konfidenz praktisch nicht erreichbar ist.

Die schwarzen Linien beziehen sich auf die Nullpunkte der jeweiligen Achsen und erlauben somit eine einfache Identifikation der 4 Quadranten. Beide vertikalen gestrichelten Linien zeigen den Schwellenwert gemäss Logib Modul 1 an.

Auf jedem Scatterplot werden auch die Korrelationskoeffizienten angezeigt, wobei sich ρ auf die Pearson-Korrelation und τ auf die Kendall-Korrelation beziehen.

Abbildung 5: Vergleich der Resultate M1 und M2



4.2.2 Unterschiedliche Vorzeichen M1 vs. M2

Fragestellung:

Wie häufig kommt es vor, dass M1 und M2 unterschiedliche Vorzeichen anzeigen (z.B. M1 zeigt Ungleichgewicht zu Ungunsten Frauen, M2 zu Ungunsten Männer)?

Wie schon aus den Scatterplots der Abbildung 5 hervorging, kommt es beim Vergleich zwischen den M1 und M2 Resultaten zu Änderungen des Vorzeichens. Die Tabelle 5 zeigt den Anteil aller Kontrollen, die je nach Modul 1 und Modul 2 eine Differenz zu Ungunsten der Frauen oder der Männer identifiziert haben. Dabei ist wichtig zu betonen, dass bei diesen Tabellen nur jeweils das Vorzeichen des Resultats berücksichtigt wurde, nicht jedoch, ob die Effekte signifikant (M1) oder über einem gewissen Wert (Modul 2) waren. Die rot eingefärbten Werte beziehen sich auf Diskrepanzen, also dass ein Modul einen Nachteil für Frauen fand, während das andere Modul einen Nachteil für Männer fand.

Tabelle 5: Anteil Kontrollen mit Vorzeichenwechsel

Umlegung FW1		
M1 / M2	Zu Ungunsten der Frauen	Zu Ungunsten der Männer
$\beta \leq 0$	61.40%	19.30%
$\beta > 0$	1.75%	17.54%
Umlegung 3a		
M1 / M2	Zu Ungunsten der Frauen	Zu Ungunsten der Männer
$\beta \leq 0$	57.89%	24.56%
$\beta > 0$	5.26%	12.28%
Umlegung 2a		
M1 / M2	Zu Ungunsten der Frauen	Zu Ungunsten der Männer
$\beta \leq 0$	57.89%	22.81%
$\beta > 0$	5.26%	14.04%
Umlegung 3b		
M1 / M2	Zu Ungunsten der Frauen	Zu Ungunsten der Männer
$\beta \leq 0$	57.89%	21.05%
$\beta > 0$	5.26%	15.79%

Je nach Art der Umlegung ändern die Werte ein wenig, sind jedoch so stabil, dass eine allgemeine Aussage gemacht werden kann:

- Zwischen 70% und 78% der Kontrollen führen zum gleichen Vorzeichen
- Bei rund 20% bis 25% der Kontrollen findet das Modul 1 einen Effekt zu Ungunsten der Frauen, während das Modul 2 einen Effekt zu Ungunsten der Männer findet.
- Bei rund 2% bis 5% der Kontrollen findet das Modul 1 einen Effekt zu Ungunsten der Männer, während das Modul 2 einen Effekt zu Ungunsten der Frauen findet.

Bei den Resultaten in der Tabelle 5 wird nur das Vorzeichen des geschätzten Resultats berücksichtigt, nicht jedoch die Grösse des Effekts oder das Resultat der Kontrolle (Grenzwertüberschreitung oder nicht). Die Tabelle 6 zeigt ein etwas detaillierteres Bild, indem bei beiden Modulen nach der Stärke des Effekts unterschieden wird⁵. Aufgrund der geringen Aussagekraft von statistischen Tests bei diesen kleinen Unternehmungen, wird bei Modul 1 nur der Koeffizient berücksichtigt, um die Resultate in 4 Kategorien zu unterteilen:

⁵Im Gegensatz zur Tabelle 5 wird hier die Anzahl Kontrollen und nicht der prozentuale Anteil angegeben. Dadurch wird die Tabelle lesbarer und aufgrund der zum Teil sehr kleinen Zahlen wären prozentuale Angaben eher irreführend.

$\beta \leq -0.05$	Ein starker Effekt zu Ungunsten der Frauen: Grenzwert überschritten (sofern signifikant).
$-0.05 < \beta \leq 0$	Effekt zu Ungunsten der Frauen: Grenzwert nicht überschritten
$0 < \beta \leq 0.05$	Effekt zu Ungunsten der Männer: Grenzwert nicht überschritten.
$0.05 < \beta$	Ein starker Effekt zu Ungunsten der Männer: Grenzwert überschritten (sofern signifikant).

Beim Modul 2 wurde eine neutrale Kategorie ($\text{Score} \leq 2.5$) und je zwei Kategorien zu Ungunsten der Frauen und der Männer definiert:

$(-\infty, -5)$	Der Score liegt über 5 und fällt zu Ungunsten der Frauen aus
$[-5, -2.5)$	Der Score liegt über 2.5, nicht aber über 5 und fällt zu Ungunsten der Frauen aus.
$[-2.5, 2.5]$	Der Score liegt unterhalb von 3 (egal in welche Richtung)
$(2.5, 5]$	Der Score liegt über 2.5, nicht aber über 5 und fällt zu Ungunsten der Männer aus.
$(5, \infty)$	Der Score liegt über 5 und fällt zu Ungunsten der Männer aus

Rote Zahlen in der Tabelle 6 heben widersprüchliche Ergebnisse zwischen M1 und M2 hervor, orangefarbene Ergebnisse weisen auf leicht widersprüchliche Ergebnisse hin. Wir definieren es als leicht widersprüchlich, wenn zum Beispiel M2 ein neutrales Ergebnis anzeigt, während M1 einen Effekt gegen eines der Geschlechter zeigt.

Widersprüchliche Ergebnisse: überschreiten der Grenzwerte mit umgekehrten Vorzeichen

Bei den vier verschiedenen Umlegungen kommt es bei einer Umlegung zu einem widersprüchlichen Resultat (Umlegung 3b): Das Modul 1 findet einen starken Effekt zu Ungunsten der Frauen, während das Modul 2 einen starken Effekt zu Ungunsten der Männer entdeckt. Bei den anderen Umlegungen kommt es bei dieser Kontrolle nur zu einem leicht widersprüchlichen Resultat, weil dort das Modul 1 keinen Effekt über 5% findet.

Unter einem Score von 5 gemäss M2, über Grenzwert gemäss M1

Neben dieser Kontrolle gibt es noch einige leicht widersprüchliche Resultate, insbesondere einige Kontrollen, bei denen das Modul 1 einen starken Effekt zu Ungunsten der Frauen findet, während das Modul 2 ein neutrales Resultat oder sogar einen kleinen Effekt zu Ungunsten der Männer findet. Bei der Umlegung 2a zum Beispiel sind es 18 der 57 Kontrollen (rund 31.5%).

Es gibt auch einige wenige Fälle (zwischen 1 und 3, je nach Umlegung), bei denen das Modul 1 einen starken Effekt zu Ungunsten der Männer findet, das Modul 2 jedoch einen neutralen Effekt oder nur einen kleinen Effekt zu Ungunsten der Männer schätzt.

Bei dieser Art von widersprüchlichen Resultaten gilt es natürlich zu beachten, dass gewisse M1-Resultate selbst bei grösseren Stichproben nicht signifikant würden und somit kein Widerspruch bestehen würde, da die Unternehmung dann gemäss beiden Modulen die Kontrolle die Kriterien der Kontrolle erfüllen würden.

Unter Grenzwert gemäss M1, über einem Score 5 gemäss M2

Den umgekehrten Fall, dass Unternehmen bei einer Kontrolle mit Modul 1 ein Resultat unterhalb des Grenzwertes haben, jedoch beim Modul 2 über dem Score von 5 liegen, gibt es auch, aber sehr selten. Zum Beispiel bei der Umlegung 3a gibt es insgesamt 5 Kontrollen (rund 8.8%) mit einem M1-Resultat unterhalb der 5%-Schwelle, aber mit einem Score über 5 im Modul 2 (4x zu Ungunsten der Frauen, 1x zu Ungunsten der Männer).

Diese detaillierte Analyse zeigt relativ klar, dass die sehr widersprüchlichen Ergebnisse praktisch nie vorkommen, es aber durchaus dazu kommen kann, dass das Modul 1 grosse geschlechtsspezifische Unterschiede findet, während das Modul 2 eher auf eine neutrale Situation hindeutet.

Tabelle 6: Anzahl Kontrollen mit Vorzeichenwechsel (detailliert)

Umlegung FW1

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5]$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	10	4	11	2	0
$-0.05 \leq \beta \leq 0$	3	4	11	1	0
$0 < \beta \leq 0.05$	0	0	3	3	1
$0.05 < \beta$	0	0	2	1	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 3a

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5]$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	9	5	12	3	0
$-0.05 \leq \beta \leq 0$	3	2	11	2	0
$0 < \beta \leq 0.05$	1	1	3	1	1
$0.05 < \beta$	0	0	1	1	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 2a

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5]$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	12	5	15	3	0
$-0.05 \leq \beta \leq 0$	0	2	7	2	0
$0 < \beta \leq 0.05$	1	1	4	2	1
$0.05 < \beta$	0	0	1	0	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 3b

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5]$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	10	5	13	2	1
$-0.05 \leq \beta \leq 0$	2	2	9	1	0
$0 < \beta \leq 0.05$	1	1	3	3	0
$0.05 < \beta$	0	0	2	1	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

4.2.3 Verteilung der Resultate nach Geschlecht

Fragestellung:

In welchem Bereich des Risikoscores bzw. des Geschlechterkoeffizienten kommen Ungleichgewichte zu Ungunsten der Männer vor?

Die Abbildung 6 zeigt die Verteilung der Resultate nach dem benachteiligten Geschlecht.



Wenn das Modell eine Differenz zu Ungunsten der Männer findet, ist dieser Effekt meistens relativ klein (1 Ausnahme), während bei Differenzen zu Ungunsten der Frauen durchaus auch grössere Koeffizienten respektive Scores gefunden werden.

4.2.4 Wo muss der Grenzwert festgelegt werden?

Fragestellung:

Bei welchem Score muss der Grenzwert in M2 (Risikostufe "sehr hoch") festgelegt werden, damit der Anteil der Unternehmen in dieser Kategorie identisch ist zum Anteil der Unternehmen, die den Grenzwert in M1 (signifikant von Null verschieden) überschreiten?

Die Frage, wo der Grenzwert festgelegt werden muss, ist von zentraler Bedeutung. In diesem Abschnitt präsentieren wir Statistiken, die bei dieser Frage in Betracht gezogen werden können. In der Abbildung 7 sehen wir die kumulative Verteilung der M2-Gesamtscores, sowie zwei horizontale Linien. Diese horizontalen Linien entsprechen der Anzahl Kontrollen, die gemäss Modul 1 unterhalb des roten Bereichs landen und somit den neuen Grenzwert nicht überschreiten würden. Wir nutzen hier zwei Kriterien:

rot Die rote Linie zeigt die Anzahl der Unternehmungen, für die kein signifikanter Effekt über 5% entdeckt wurde. Für die Umlegung FW1 sehen wir, dass rund 43 Unternehmungen unterhalb dieses kritischen Werts liegen, während rund 10 Firmen im roten Bereich zu liegen kommen.

grau Da der statistische Test in dieser Analyse sehr fraglich ist (wenige Datenpunkte), zeigt die graue Linie die Anzahl der Firmen mit einem Geschlechtskoeffizienten unter 5%, egal ob signifikant oder nicht.

Zwei Kriterien zu nutzen macht Sinn, weil die rote Linie die Anzahl Unternehmungen im roten Bereich eher unterschätzt (fehlende Anzahl Datenpunkte), während die graue Linie eher ein oberes Limit der Anzahl Unternehmungen angibt.

Wenn das Ziel sein soll, die gleiche Anzahl Firmen im roten (M1) Bereich und in der Kategorie "Sehr hoch" (M2) zu haben, dann kann man sich auf den Schnittpunkt der horizontalen Linien mit der kumulativen Verteilung der M2 konzentrieren. Somit müsste der Grenzwert je nach Umlegung und Kriterium (horizontale Linie) irgendwo zwischen 2 und 5.5 zu liegen kommen.

Neben der gleichen Anzahl Unternehmungen könnte man aber auch ein etwas detaillierteres Kriterium benutzen. Grundsätzlich wäre es wünschenswert, dass beide Module für eine und dieselbe Unternehmung das gleiche Resultat (über vs. unter Grenzwert) angeben. Die Abbildung 8 zeigt den Anteil Unternehmungen, für welche beide Module zum gleichen Resultat kommen, in Funktion des Grenzwerts bei Modul 2. Wie schon zuvor haben wir auch hier 2 Kriterien beim Modul 1 benutzt.

Im oberen Teil der Grafik sieht man, dass bei einem Grenzwert von 6 ein lokales Maximum erreicht wird. Bei gewissen Umlegungen wird dieses auch noch bei sehr viel höheren Grenzwerten erreicht oder leicht übertroffen. Im unteren Teil der Grafik (Kriterium: >5%, egal ob signifikant) ist das Maximum rund um den Wert 6 sehr stabil (alle Umlegungen) und global. Wenn also ein Grenzwert von rund 6 benutzt wird, erreichen wir die höchste Kohärenz zwischen den Resultaten M1 und M2.

Abbildung 7: Kumulative Verteilung der M2 Kontrollen

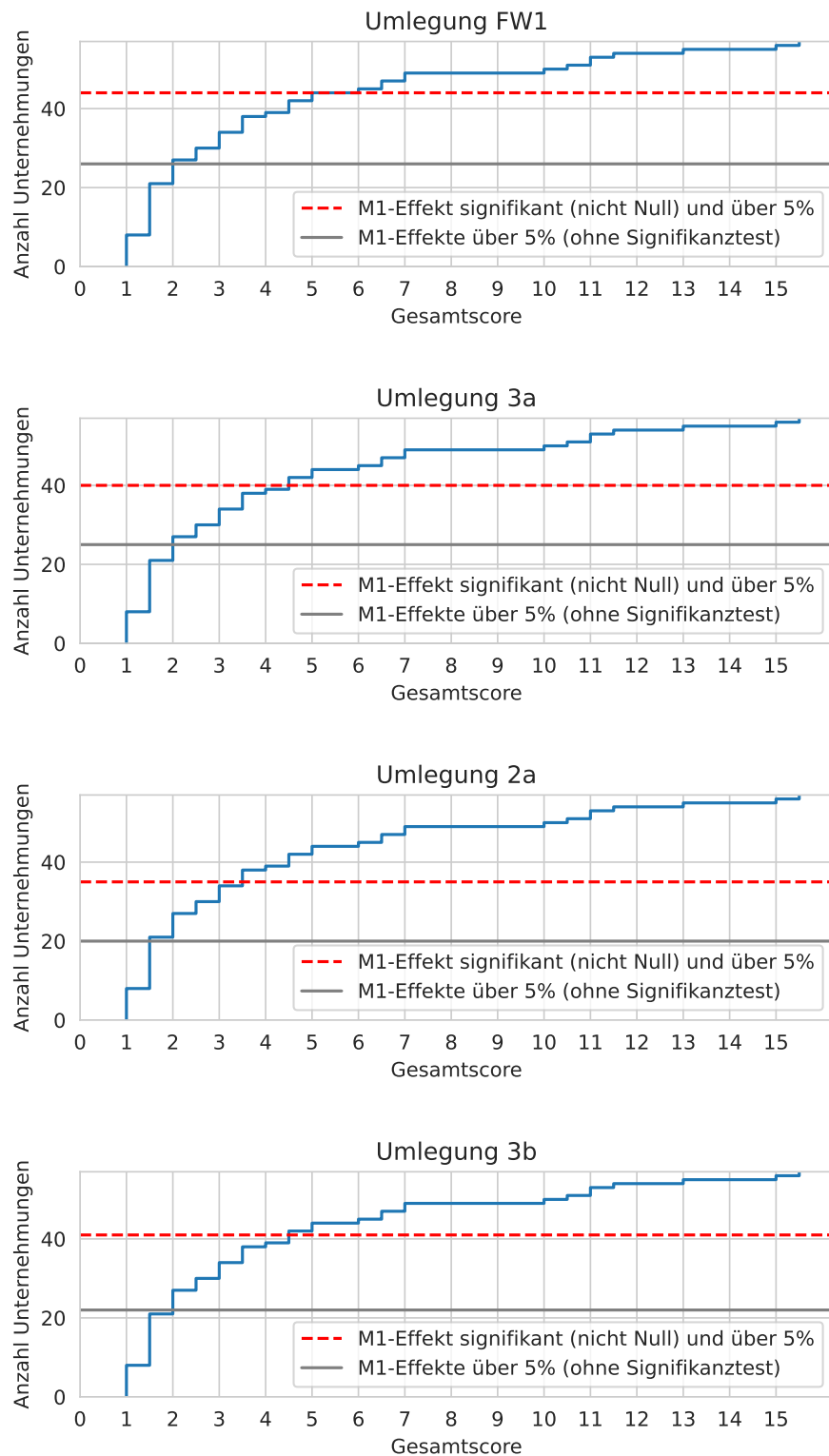
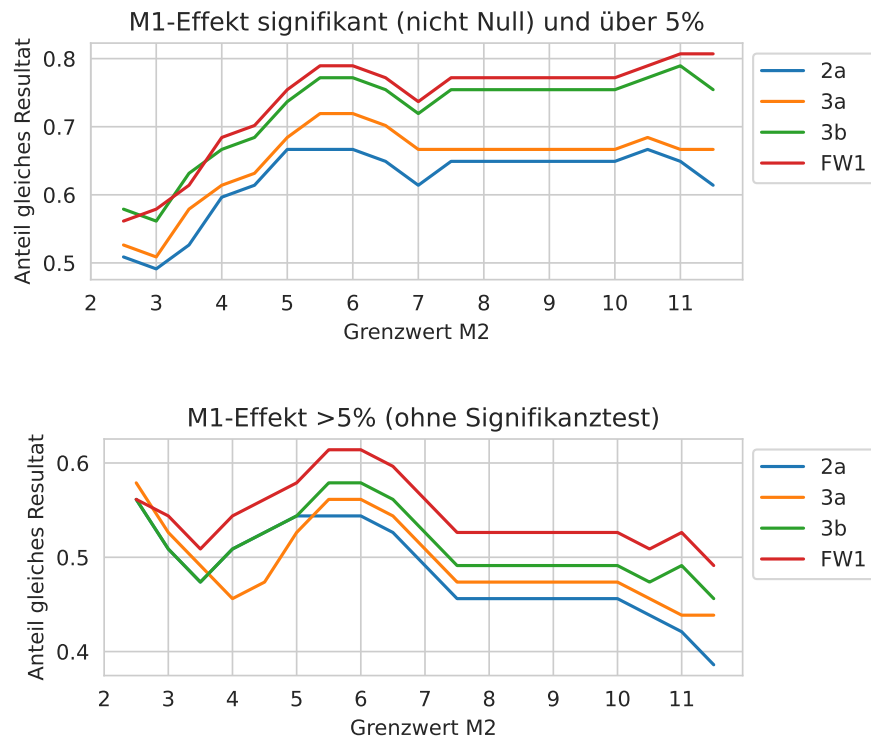


Abbildung 8: Anteil der Kontrollen mit gleichem Resultat M1 und M2



5 Leave-one-out Ansatz

In diesem Kapitel wird nun ein Vorschlag des EBG zur Reduktion der Volatilität der Ergebnisse des Moduls 2 vorgestellt. Damit soll verhindert werden, dass einzelne Personen innerhalb eines Unternehmens das Ergebnis zu stark dominieren. Diese Befürchtung wurde von Fachpersonen Lohnungleichheit im Rahmen der Besprechungen zu dieser Kalibrierung geäußert.

Die Idee und ihre Umsetzung in dieser Studie werden im Abschnitt 5.1 beschrieben. Im Abschnitt 5.2 werden deskriptive Statistiken zum Ansatz präsentiert. Diese basieren auf allen Modul 2 Kontrollen, die für die Studie zur Verfügung standen, also auch auf denen, die in den vorherigen Grafiken ausgeschlossen wurden. Im Abschnitt 5.3 werden die Ergebnisse des Abschnitts 4.2 noch einmal unter Berücksichtigung des Leave-one-out-Ansatzes dargestellt.

5.1 Idee und Umsetzung

Um zu vermeiden, dass die Ergebnisse einer Modul-2-Analyse auf einer einzelnen Person innerhalb des Unternehmens beruhen, wird beim Leave-one-out-Ansatz ein alternativer Score berechnet, bei dem jeweils eine Person aus der Berechnung des Scores ausgeschlossen wird. Es ist wichtig zu erwähnen, dass die Schätzung nicht neu durchgeführt wird, sondern nur die Berechnung des Scores. Da man *a priori* nicht wissen kann, welche Person das Ergebnis am meisten beeinflusst, wird dieser alternative Score für alle Personen berechnet, die in den Risikopaaren vorkommen (Personen mit tieferem Lohn als erwartet und Vergleichspersonen). Dies führt zur Berechnung einer potenziell sehr grossen Anzahl alternativer Scores. Der Leave-one-out-Score (kurz LOO-Score) entspricht dann dem kleinsten absoluten Wert aller berechneten Scores (inklusive des Standardscores).

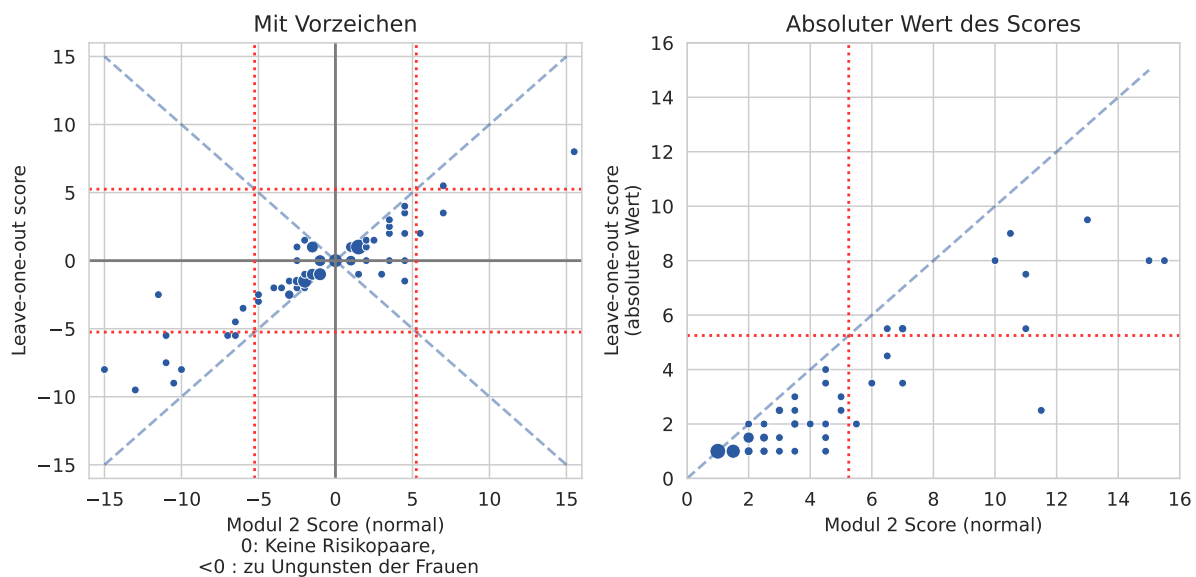
Für die vorliegende Studie konnte dieser LOO-Score direkt aus den Daten der Detailergebnisse (Excel-Datei) der jeweiligen Kontrollen berechnet werden. Um sicherzustellen, dass die Berechnung des Scores korrekt ist, wurde auch der Standardscore mit dem gleichen Skript berechnet und mit den offiziellen Scores von Logib

verglichen. Diese Überprüfung ergab, dass die Berechnung korrekt ist.

5.2 Deskriptive Statistik

Bevor die Auswirkungen des LOO-Ansatzes auf die Kalibrierung diskutiert werden, ist es sinnvoll, die Ergebnisse des LOO-Ansatzes deskriptiv zu betrachten. Die Abbildung 9 vergleicht den normalen Score mit dem LOO-Score⁶. Auf der linken Seite sind beide Scores mit Vorzeichen dargestellt, wobei negative Werte einer Situation zu Ungunsten der Frauen entsprechen. Auf der rechten Seite sind die gleichen Daten als absolute Werte dargestellt. Aufgrund der begrenzten Anzahl möglicher Werte kommt es in dieser Grafik zu Überlappungen verschiedener Punkte. Daher sind die Punkte mit mehr Beobachtungen etwas grösser, um sie optisch hervorzuheben.

Abbildung 9: Vergleich des normalen Scores mit dem LOO-Scores



Aufgrund der Definition des LOO-Scores sind dessen Werte immer mindestens so nah am horizontalen Ursprung (0) wie am vertikalen Ursprung. Es ist relativ einfach zu sehen, dass in den allermeisten Fällen die LOO-Scores näher bei null liegen als die normalen Scores. Die roten Linien beziehen sich auf einen Wert von 5.25 und somit sind alle Punkte unterhalb der Linie im Bereich, in welcher der Grenzwert⁷ nicht überschritten würde.

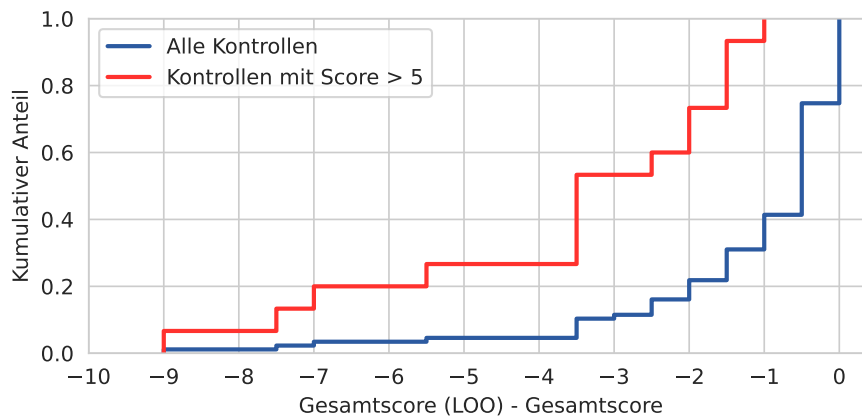
Von den 88 Kontrollen scheint es 5 zu geben, die gemäss dem Standardscore den Grenzwert überschreiten, nicht jedoch mit dem LOO-Score. Drei dieser 5 Punkte befinden sich in negativen (zu Ungunsten der Frauen) und 2 im positiven (zu Ungunsten der Männer) Bereich. Ganz allgemein kann man sehen, dass die Scores deutlich nach null korrigiert werden.

Die Abbildung 10 verdeutlicht dies ziemlich klar. Auf der horizontalen Achse wird der Veränderung des Scores vom normalen auf den LOO-Score angezeigt und auf der vertikalen Achse die kumulative Proportion. Die blaue Kurve zeigt die Verteilung aller Kontrollen, während die rote Kurve nur die Verteilung bei Kontrollen mit einem normalen Score über 5 zeigt.

⁶Um die Grafik aussagekräftiger zu machen, wurden die Scores ohne Risikopaare auf null gesetzt, obwohl Logib in diesem Fall normalerweise den Wert 1 ausgibt. Hier ist es jedoch sinnvoll, diesen Wert auf null zu setzen, um zu zeigen, dass durch das Weglassen von Personen teilweise keine Risikopaare mehr vorhanden sind.

⁷Diese Aussage bezieht sich auf die Überlegung des EBG während der Erstellung dieser Studie, den neuen Grenzwert bei 5 festzulegen. Kontrollen mit einem Wert bis 5 würden den Grenzwert nicht überschreiten, ab 5.5 läge eine Überschreitung vor.

Abbildung 10: Verteilung des LOO-Gewinns



Betrachtet man die Grafik von links her, so zeigt sich, dass ca. 10% aller Fälle eine Reduktion von 3.5 oder mehr Punkten aufweisen. Betrachtet man die Grafik auf der rechten Seite, so zeigt sich, dass fast 60% der Fälle keine Reduktion oder nur eine Reduktion von 0.5 aufweisen. Dieser hohe Anteil ist natürlich auch auf die vielen Kontrollen mit sehr niedrigen Scores zurückzuführen, bei denen eine weitere Reduktion schwierig bis unmöglich ist. Die rote Kurve zeigt, dass alle Normalscores über 5 durch den LOO-Ansatz um mindestens einen Punkt reduziert werden. Bei etwa der Hälfte der Kontrollen beträgt die Reduktion 3.5 Punkte oder mehr.

Die Abbildung 11 gibt ein noch viel detaillierteres Bild. Auf der vertikalen Achse sind alle 88 Kontrollen nach dem normalen Score sortiert. Die grünen Punkte beziehen sich auf den normalen Score von Logib, die blauen Kreuze auf den kleinsten LOO-Score und die orangen Rhomben zeigen den grössten Wert aller LOO-Scores an.

Es gibt verschiedene spannende Aussagen aus dieser Grafik:

- Bei etwa 20 bis 25 der 88 Kontrollen gibt es deutliche Abweichungen nach links oder rechts. Das Weglassen einer einzigen Person führt in diesen Fällen zu deutlichen Veränderungen des Scores.
- Bei Kontrollen mit hohen Scores tendieren die Verschiebungen eher in Richtung einer Verstärkung des Effekts, d.h. bei Kontrollen mit einem Ergebnis zu Ungunsten der Frauen gibt es LOO-Scores, die ein noch viel stärkeres Ergebnis zu Ungunsten der Frauen finden (unterer Teil der Grafik). Dasselbe gilt für die Ergebnisse zu Ungunsten der Männer (oberer Teil).
- Es gibt verschiedene Fälle, in denen der niedrigste und der höchste Wert nicht das gleiche Vorzeichen haben (mehr dazu in der Tabelle 7).

In der Tabelle 7 werden einige Statistiken aus der Grafik 11 abgeleitet. Dabei ist zu beachten, dass sich die Prozentwerte nicht zu 100% addieren, weil sie sich zum Teil überschneiden. Diese Werte dienen in erster Linie einem besseren Verständnis des LOO-Scores, sind aber für die Umsetzung des Scores im Modul 2 nicht direkt relevant.

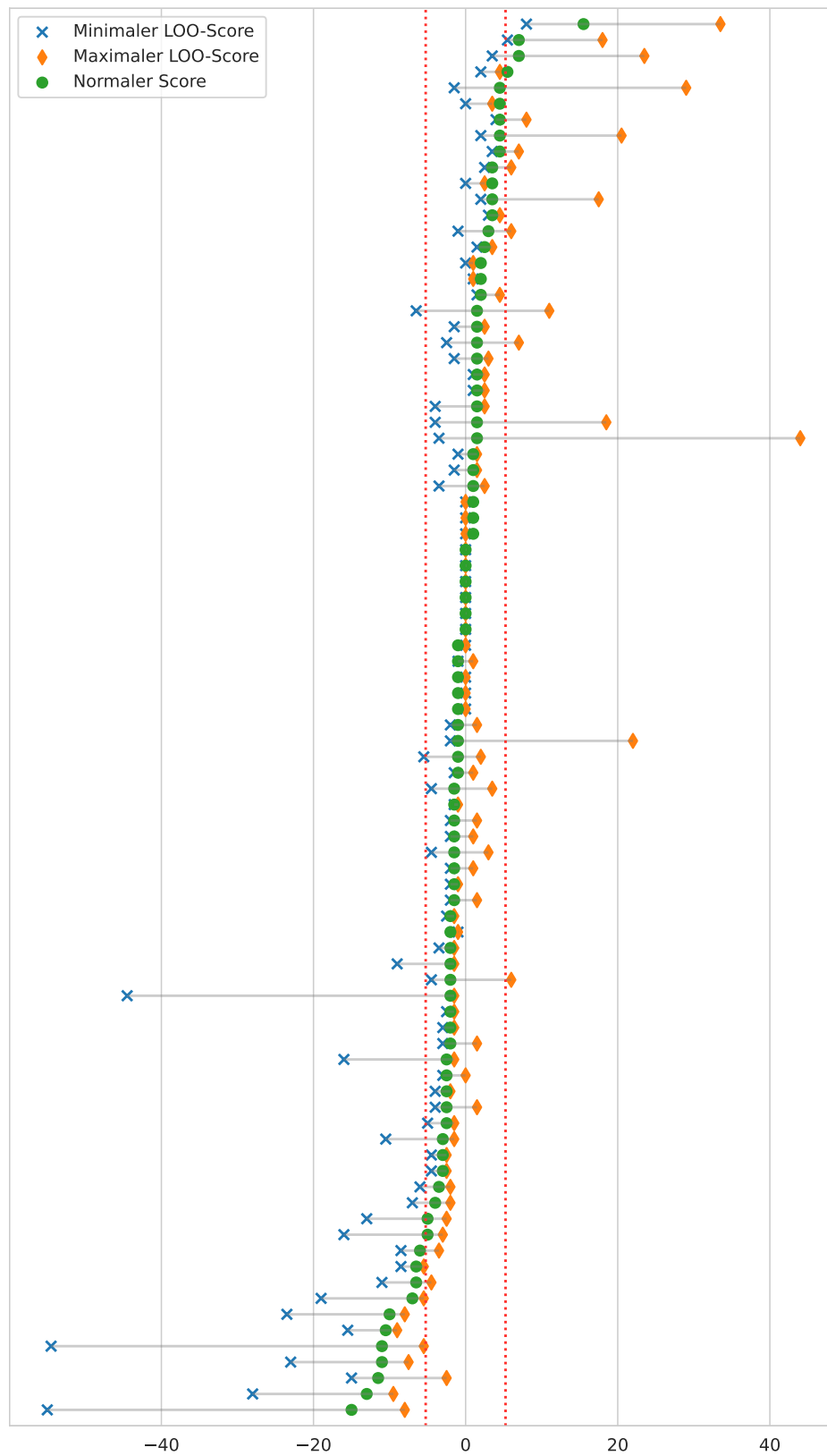
Tabelle 7: Deskriptive Statistik zum Leave-one-out-Ansatz

Kein Vorzeichenwechsel	70.45%
Kein Vorzeichenwechsel und keine Werte über 5	26.14%
Wechsel des Vorzeichens	29.55%
Wechsel des Vorzeichens und mindestens auf einer Seite über 5	10.23%
Wechsel des Vorzeichens und auf beiden Seiten über 5	1.14%

Bemerkung: Anteile basieren auf 88 Kontrollen

Bei 70.45% der Fälle kommt es zu keinem Vorzeichenwechsel, wenn man den LOO-Ansatz nutzt. Bei 26.14%

Abbildung 11: Minimale und maximale LOO-Scores



aller Fälle gibt es keinen Vorzeichenwechsel und sämtliche LOO-Scores sind unterhalb des Grenzwertes. In diesen Fällen gibt jede Kombination das genau gleiche qualitative Resultat.

In 29.55% der Fälle ändert sich dagegen das Vorzeichen durch den Wegfall von Personen. Dabei handelt es sich jedoch häufig um sehr kleine Verschiebungen innerhalb des Bereichs, in dem die Betriebe den Grenzwert nicht überschreiten (-5 bis 5). In rund einem Drittel dieser Fälle, d.h. bei 10.23% aller Kontrollen, kommt es zu einem Vorzeichenwechsel mit mindestens einem Score über dem Grenzwert. Bei 1.14% (1 Kontrolle in der Stichprobe) kommt es sogar zu einer Situation, in der mindestens eine Grenzwertüberschreitung sowohl im negativen (zu Ungunsten der Frauen) als auch im positiven (zu Ungunsten der Männer) Bereich gefunden wird. Dieser Fall ist insofern besonders, als der normale Score keine Hinweise auf ein Problem gibt, während die LOO-Scores gegensätzliche Ergebnisse anzeigen. Da in diesem Fall der normale Score verwendet wird, würde das Unternehmen den Grenzwert nicht überschreiten.

5.3 Resultate zu den spezifischen Fragen mit dem Leave-one-out Ansatz

In diesem Abschnitt werden nun die gleichen Grafiken und Tabellen wie im Abschnitt 4.2 präsentiert, jedoch immer mit dem LOO-Score. Da die Grafiken im Abschnitt 4.2 bereits ausführlich besprochen wurden, werden hier nur einige Anmerkungen zu den Unterschieden gemacht.

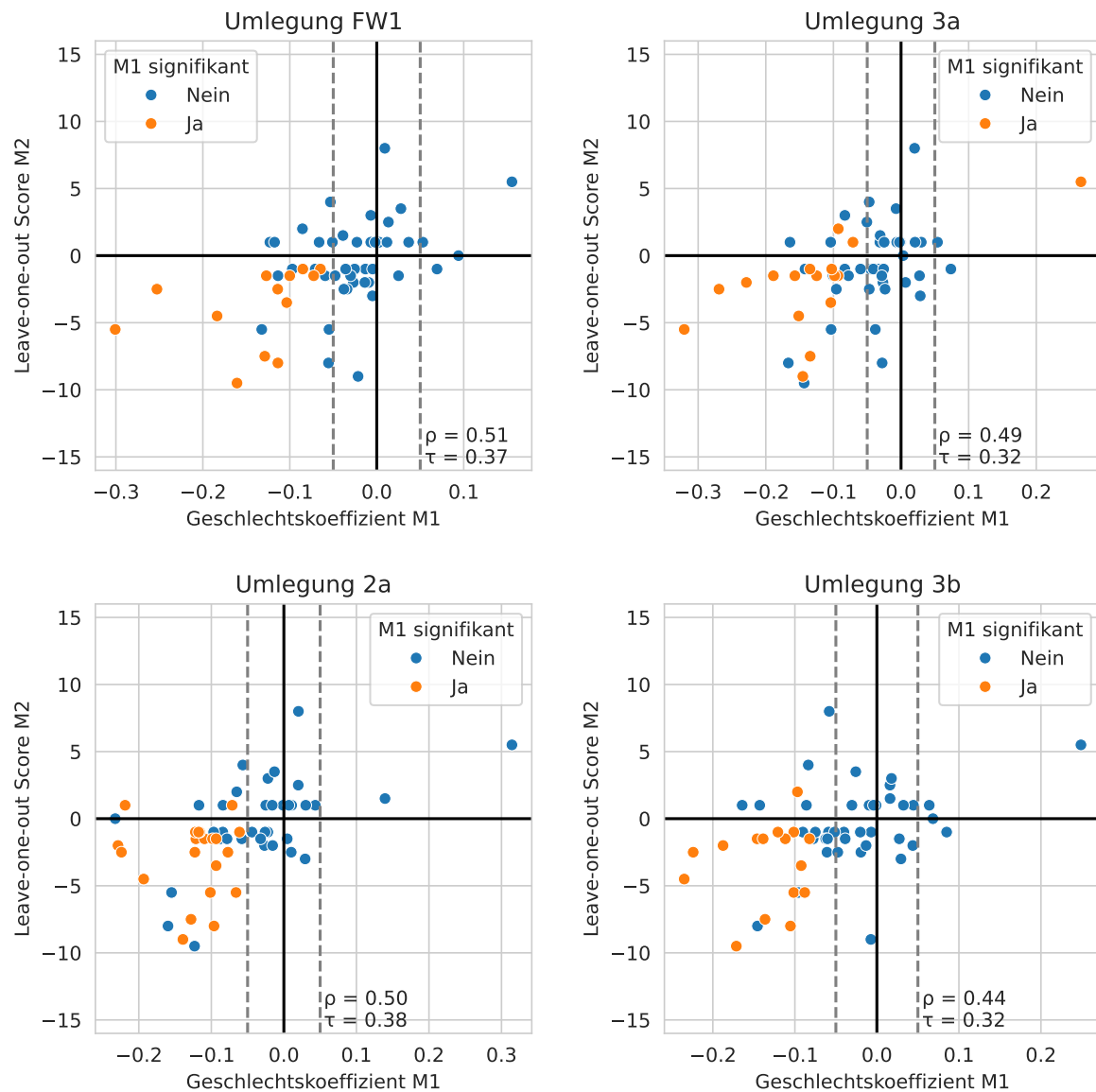
5.3.1 Vergleich der Resultate M1 vs. M2: Scatterplot

Fragestellung:

Gibt es einen Zusammenhang zwischen dem Geschlechterkoeffizienten M1 und dem Risikoscore M2?

Die Abbildung 12 entspricht der Abbildung 5 im Abschnitt 4.2.1. Die Ergebnisse sind grundsätzlich ähnlich, allerdings ist eine geringere Korrelation zwischen den Modul-1-Ergebnissen (Umlegungen) und den Modul-2-Ergebnissen zu erkennen. Dieser Rückgang ist jedoch einerseits relativ logisch und sollte andererseits nicht überbewertet werden. Die Modul 1-Werte beziehen sich auf alle Mitarbeitenden, während die LOO-Werte jeweils mit einer Person weniger berechnet wurden. Da es sich dabei typischerweise um mögliche Ausreisser handelt, würde auch das M1-Ergebnis ohne diese Person wahrscheinlich etwas anders ausfallen. Daher ist es logisch, dass die Korrelation etwas abnimmt.

Abbildung 12: Vergleich der Resultate M1 und M2 (LOO-Ansatz)



5.3.2 Unterschiedliche Vorzeichen M1 vs. M2

Fragestellung:

Wie häufig kommt es vor, dass M1 und M2 unterschiedliche Vorzeichen anzeigen (z.B. M1 zeigt Ungleichgewicht zu Ungunsten Frauen, M2 zu Ungunsten Männer)?

Die Tabelle 8 entspricht der Tabelle 6 im Abschnitt 4.2.2 und zeigt die Verteilung nach möglichen Entscheidungen bei einer Kontrolle gemäss Modul 1 und Modul 2 auf.

Rote Zahlen in der Tabelle 8 heben widersprüchliche Ergebnisse zwischen M1 und M2 hervor, orangefarbene Ergebnisse weisen auf leicht widersprüchliche Ergebnisse hin. Wir definieren es als leicht widersprüchlich, wenn zum Beispiel M2 ein neutrales Ergebnis anzeigt, während M1 einen Effekt gegen eines der Geschlechter zeigt.

Widersprüchliche Ergebnisse: überschreiten der Grenzwerte mit umgekehrten Vorzeichen

Bei den vier verschiedenen Umlegungen kommt es bei einer Umlegung zu einem widersprüchlichen Resultat (Umlegung 3b): Das Modul 1 findet einen starken Effekt zu Ungunsten der Frauen, während das Modul 2 einen starken Effekt zu Ungunsten der Männer entdeckt. Dieser Fall ist identisch wie in der Tabelle 6.

Unter einem Score von 5 gemäss M2, über Grenzwert gemäss M1

Neben dieser Kontrolle gibt es noch einige leicht widersprüchliche Resultate, insbesondere einige Kontrollen, bei denen das Modul 1 einen starken Effekt zu Ungunsten der Frauen findet, während das Modul 2 ein neutrales Resultat oder sogar einen kleinen Effekt zu Ungunsten der Männer findet. Der Anteil dieser leichten Widersprüche ist beim LOO-Score etwas grösser als beim normalen Score. Bei der Umlegung 2a zum Beispiel waren es 18 Fälle mit dem normalen Score und mit dem LOO-Score sind es 22.

Es gibt auch einige wenige Fälle (zwischen 1 und 3, je nach Umlegung), bei denen das Modul 1 einen starken Effekt zu Ungunsten der Männer findet, das Modul 2 jedoch einen neutralen Effekt oder nur einen kleinen Effekt zu Ungunsten der Männer schätzt. Dieses Resultat ist ziemlich ähnlich wie beim normalen Score.

Unter Grenzwert gemäss M1, über einem Score 5 gemäss M2

Den umgekehrten Fall, dass Unternehmen bei einer Kontrolle mit Modul 1 ein Resultat unterhalb des Grenzwertes haben, jedoch beim Modul 2 über dem Score von 5 liegen, gibt es noch seltener als mit dem Normalscore. Zum Beispiel bei der Umlegung 3a gibt es insgesamt 3 Kontrollen mit einem M1-Resultat unterhalb der 5%-Schwelle, aber mit einem Score über 5 im Modul 2 (2x zu Ungunsten der Frauen, 1x zu Ungunsten der Männer). Mit dem normalen Score waren es 5 Fälle.

Diese detaillierte Analyse zeigt relativ klar, dass die sehr widersprüchlichen Ergebnisse praktisch nie vorkommen, es aber durchaus dazu kommen kann, dass das Modul 1 grosse geschlechtsspezifische Unterschiede findet, während das Modul 2 eher auf eine neutrale Situation hindeutet. Dieser Befund wird durch die LOO-Scores eher noch verschärft.

Tabelle 8: Anzahl Kontrollen mit Vorzeichenwechsel (detailliert), LOO

Umlegung FW1

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	7	4	15	1	0
$-0.05 \leq \beta \leq 0$	1	3	14	1	0
$0 < \beta \leq 0.05$	0	0	5	1	1
$0.05 < \beta$	0	0	3	0	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 3a

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	6	4	18	1	0
$-0.05 \leq \beta \leq 0$	2	2	12	2	0
$0 < \beta \leq 0.05$	0	1	5	0	1
$0.05 < \beta$	0	0	2	0	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 2a

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	8	5	21	1	0
$-0.05 \leq \beta \leq 0$	0	0	9	2	0
$0 < \beta \leq 0.05$	0	2	6	0	1
$0.05 < \beta$	0	0	1	0	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 3b

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	7	4	18	1	1
$-0.05 \leq \beta \leq 0$	1	2	10	1	0
$0 < \beta \leq 0.05$	0	1	6	1	0
$0.05 < \beta$	0	0	3	0	1

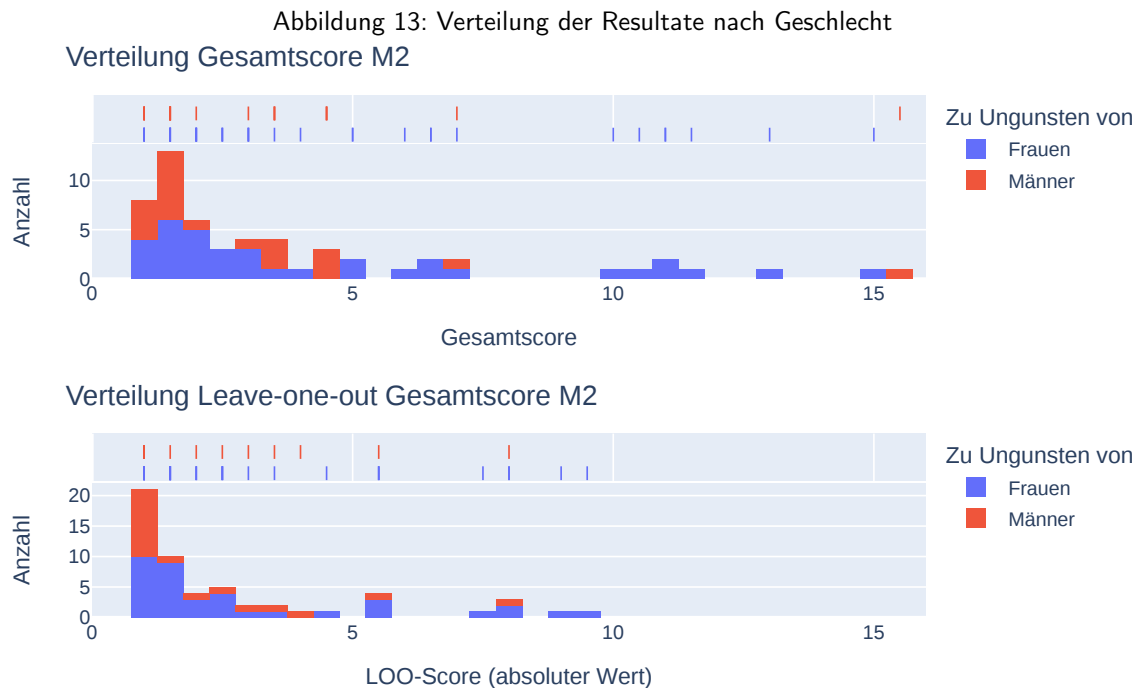
Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

5.3.3 Verteilung der Resultate nach Geschlecht

Fragestellung:

In welchem Bereich des Risikoscores bzw. des Geschlechterkoeffizienten kommen Ungleichgewichte zu Ungunsten der Männer vor?

Die Abbildung 13 zeigt die Verteilung der Ergebnisse mit dem Normalscore und dem LOO-Score nach benachteiligtem Geschlecht. Die Verteilung des LOO-Scores liegt definitionsgemäss näher bei null, da der LOO-Score im Vergleich zum normalen Score nur kleiner werden oder gleich bleiben kann.



Beim LOO-Score gibt es in dieser Stichprobe keine Werte über 10 und die grosse Mehrheit der Werte liegt unterhalb von 5.

5.3.4 Wo muss der Grenzwert festgelegt werden?

Fragestellung:

Bei welchem Score muss der Grenzwert in M2 (Risikostufe "sehr hoch") festgelegt werden, damit der Anteil der Unternehmen in dieser Kategorie identisch ist zum Anteil der Unternehmen, die den Grenzwert in M1 (signifikant von Null verschieden) überschreiten?

Die Abbildung 14 entspricht der Abbildung 7 im Abschnitt 4.2.4 und zeigt die kumulative Verteilung der Ergebnisse von Modul 2 sowie die Anzahl der Unternehmen, die den Grenzwert gemäss Modul 1 nicht überschreiten würden. Der Schnittpunkt der kumulativen Verteilung mit den horizontalen Linien wäre somit ein Grenzwert, der zu einer gleichen Anzahl von Kontrollen mit Grenzwertüberschreitung führen würde.

Im Vergleich zu den Ergebnissen in Abbildung 7 liegen diese Schnittpunkte hier deutlich niedriger, typischerweise zwischen 1.5 und 3.0. Diese Reduktion ist eine direkte Folge des LOO-Ansatzes, da die Modul-2 Scores kleiner werden. Um die gleiche Anzahl von Kontrollen mit Grenzwertüberschreitungen zu erhalten, müsste der Grenzwert noch niedriger angesetzt werden.

Die Abbildung 15 entspricht der Abbildung 8 und zeigt den Anteil der Unternehmen, für die beide Module zum gleichen Ergebnis führen, in Abhängigkeit vom Grenzwert in Modul 2 und unter Berücksichtigung des LOO-Ansatzes. Wie zuvor wurden auch hier 2 Kriterien für Modul 1 verwendet.

Abbildung 14: Kumulative Verteilung der M2 Kontrollen

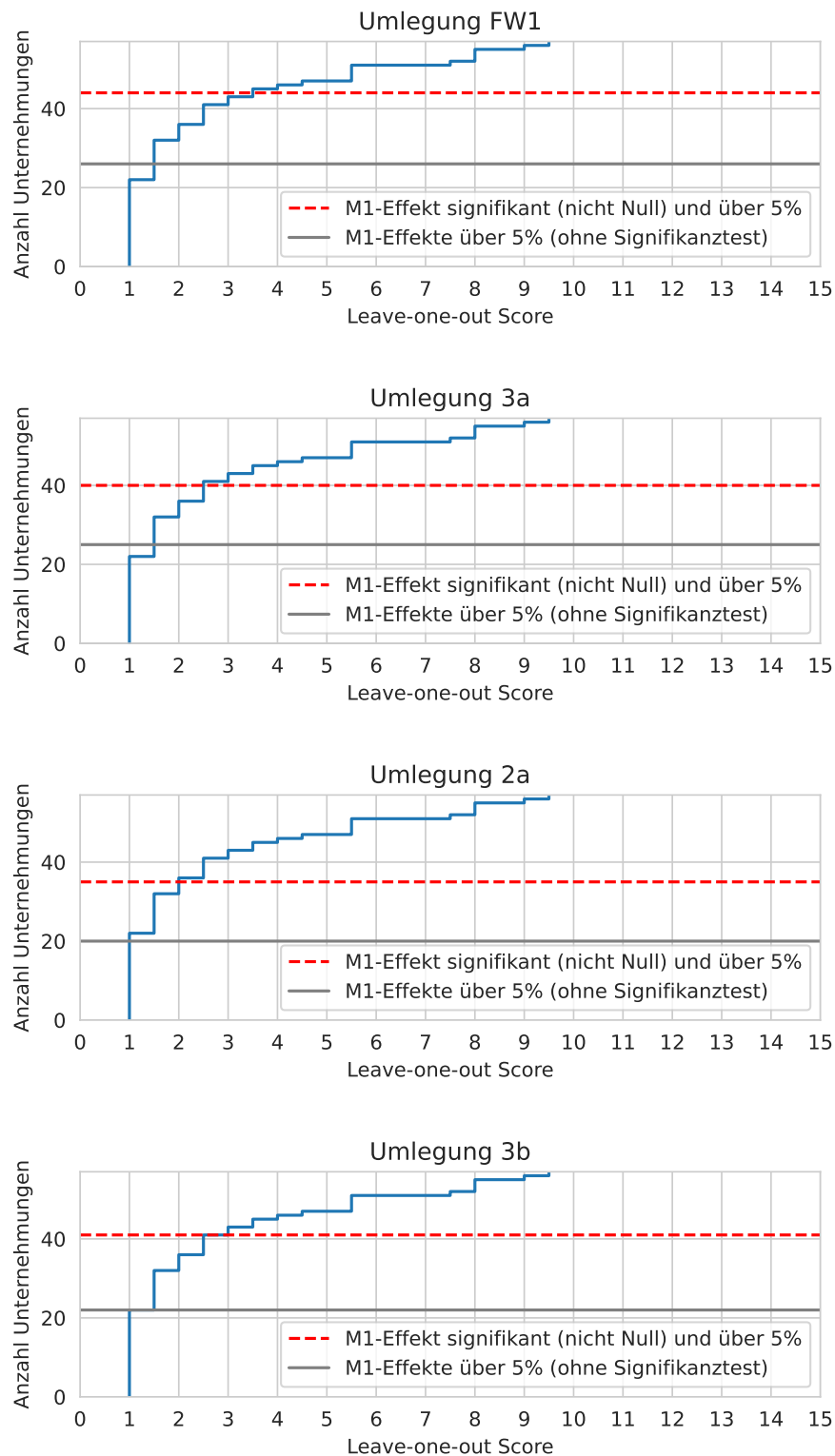
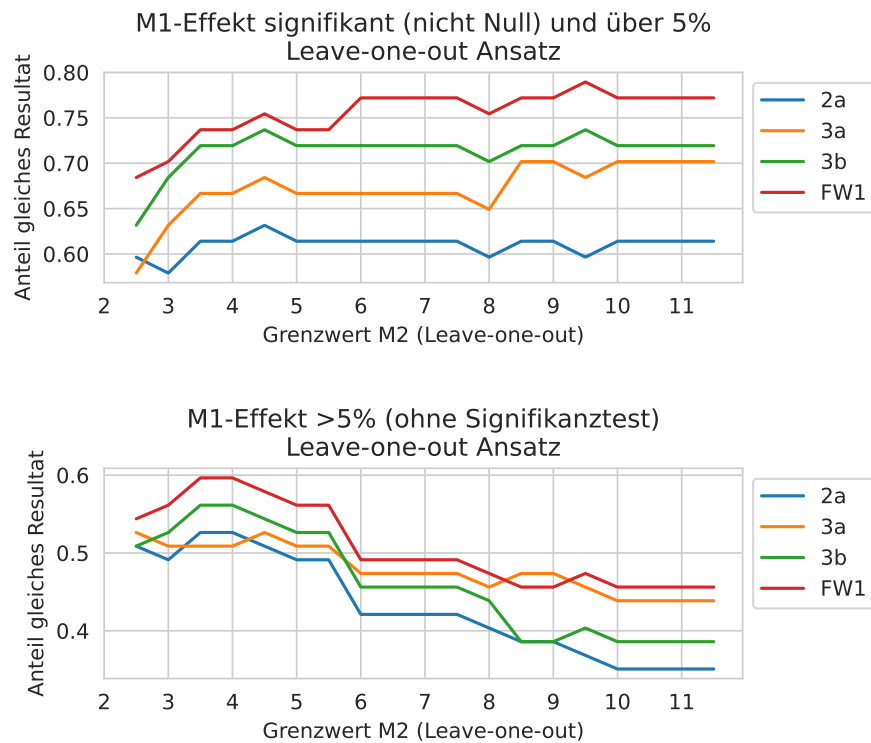


Abbildung 15: Anteil der Kontrollen mit gleichem Resultat M1 und M2



Im Vergleich zur Abbildung 8 ist der Anteil der gleichen Ergebnisse bei niedrigen Grenzwerten tendenziell etwas höher, bei höheren Grenzwerten hingegen niedriger. Bei den LOO-Werten gibt es zumindest in der oberen Grafik keine eindeutige Aussage, wo der Grenzwert liegen sollte. Je nach Umlegung ergeben sich andere Maxima und generell ändert sich der Anteil nicht allzu sehr in Abhängigkeit vom Grenzwert.

Im unteren Teil der Abbildung findet sich ein Maximum zwischen 3.5 und 4.0, was einem etwas niedrigeren Wert als in Abbildung 8 entspricht. Wie bereits erwähnt, führt die Einführung des LOO-Ansatzes zu einer Verringerung der Scores, sodass ein noch niedrigerer Grenzwert festgelegt werden müsste, um möglichst ähnliche Ergebnisse wie in Modul 1 zu erhalten.

6 Schlussbemerkungen

Die statistische Auswertung der 57 verfügbaren Modul-2-Kontrollen hat gezeigt, dass die Korrelation zwischen den Scores von Modul 2 und dem Modul-1-Koeffizienten je nach Umlegung zwischen 0.45 und 0.60 liegt, wobei die Pearson-Korrelation typischerweise etwas höher ist als die rangbasierte Kendall-Korrelation.

Die Einführung eines Leave-one-out-Scores (LOO-Scores) führt zu rund 0.03 bis 0.05 tieferen Korrelationen zwischen den M1 und M2 Resultaten.

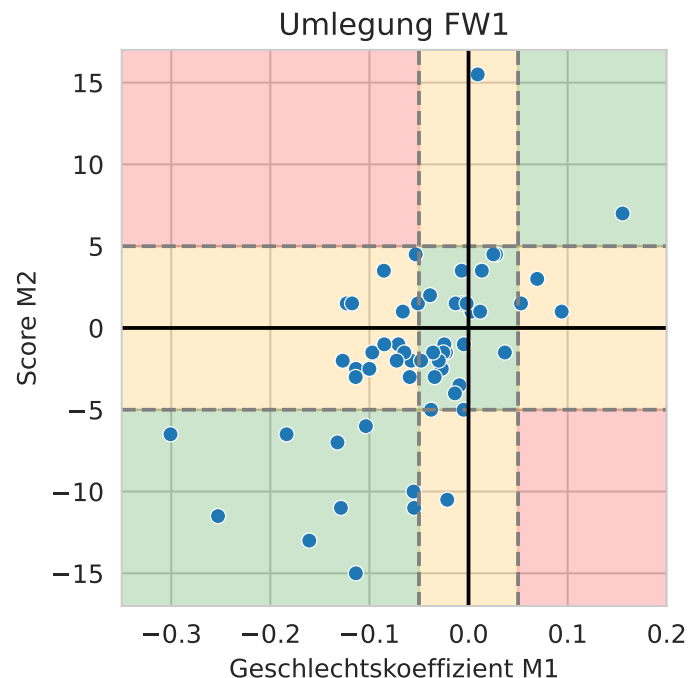
Trotz dieser nicht sehr starken Korrelation gibt es nur wenige Unternehmen, bei denen die beiden Module stark widersprüchliche Ergebnisse liefern würden. Nur bei einer der vier betrachteten Umlegungen kommt es zu einer Situation, in der das Unternehmen nach beiden Modulen den Grenzwert überschreiten würde, allerdings mit umgekehrten Vorzeichen. Dieser Fall ergibt sich sowohl beim normalen Score als auch beim LOO-Score. Weitaus häufiger ist die Situation, dass Unternehmen nach einem Modul den Grenzwert einhält und nach dem anderen diesen jedoch überschreitet. In den meisten dieser Fälle würden die Unternehmen nach M2 den Grenzwert einhalten, nach M1 jedoch nicht. Der LOO-Ansatz führt zu noch mehr Fällen, in welchen das Modul 1 auf ein Problem hinweist, während das Modul 2 keine Grenzwertüberschreitung findet.

Die Frage nach dem richtigen Grenzwert bei Modul 2 kann nicht eindeutig beantwortet werden. Wenn die Anzahl Unternehmen, die den Grenzwert einhalten, gleich hoch sein soll, dann **müsste der Grenzwert irgendwo zwischen 2 und 5.5 liegen**. Wenn man jedoch versucht, die Anzahl widersprüchlicher Ergebnisse zu minimieren, dann wäre **eher ein Grenzwert zwischen 5 und 6 zu berücksichtigen**.

Bei Einführung des LOO-Ansatzes müsste der Grenzwert nach beiden Kriterien noch etwas niedriger angesetzt werden. Wenn die Anzahl der Kontrollen mit Grenzwertüberschreitung als Kriterium verwendet werden soll, müsste der Grenzwert zwischen 1.5 und 3 liegen. Um möglichst ähnliche Ergebnisse zwischen M1 und M2 zu erhalten, sollte ein Grenzwert zwischen 3.5 und 4.0 in Betracht gezogen werden.

Die Abbildung 16 zeigt am Beispiel der Umlegung FW1 und einem Grenzwert 5 für das Modul 2 (normaler Score), wie sich die Situation präsentieren würde.

Abbildung 16: Hauptresultat am Beispiel der Umlegung FW1



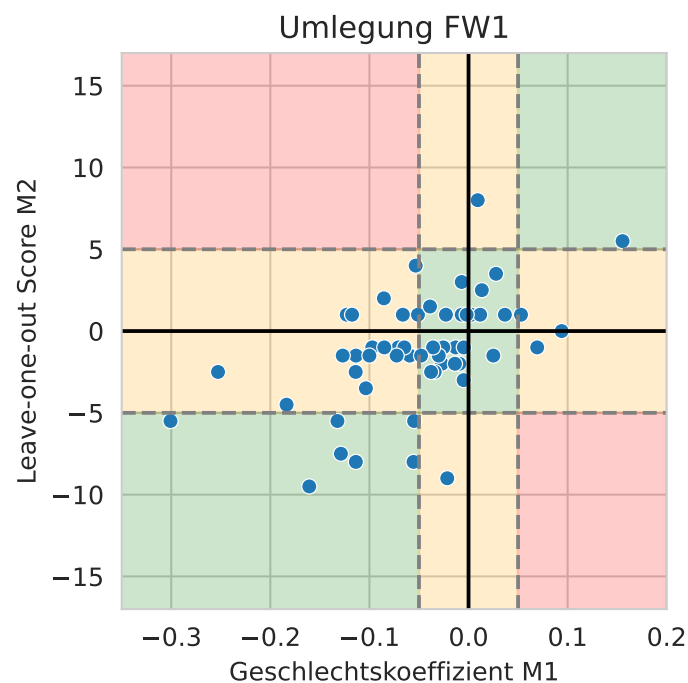
Alle Punkte, die auf einer grün hinterlegten Fläche liegen, beziehen sich auf eine Kontrolle, die bei beiden

Modulen das gleiche qualitative Resultat erhalten würde. Die gelb hinterlegten Flächen zeigen Zonen, wo es zu leichten Widersprüchen kommt, da die Unternehmung nur gemäss eines der beiden Module den Grenzwert einhalten würde. Dieser Fall ist natürlich insofern problematisch, als die beiden Module unterschiedliche qualitative Resultate liefern.

Die rot hinterlegten Zonen wären die problematischsten Fälle, da sich beide Module zwar einig sind, dass die Grenzwerte überschritten werden, jedoch unter widersprüchlichen Vorzeichen.

Die gleiche Grafik mit dem LOO-Score findet sich in Abbildung 17. Grundsätzlich ist die Grafik sehr ähnlich, aber es ist deutlich zu erkennen, dass die Punkte auf der vertikalen Achse näher am Nullpunkt liegen. Insbesondere im linken unteren Quadranten ist eine vertikale Verschiebung der Punkte vom grünen Bereich (nach beiden Modulen über dem Grenzwert) in den gelben Bereich zu erkennen, was bedeutet, dass eine Kontrolle nach Modul 1 den Grenzwert überschreitet, nach Modul 2 jedoch nicht. Sowohl beim normalen als auch beim LOO-Score gibt es bei dieser Verschiebung keine Punkte im roten Bereich, d.h. Unternehmen, die je nach Modul mit unterschiedlichem Vorzeichen den Grenzwert überschreiten.

Abbildung 17: Hauptresultat am Beispiel der Umlegung FW1 (Leave-one-out)



Literatur

- EBG**, "Anpassung der Toleranzschwelle 2024," Eidgenössisches Büro für die Gleichstellung von Frau und Mann EBG. Online verfügbar auf <https://www.ebg.admin.ch/de/lohngleichheit-analysieren-mit-logib> oder zum direkten Download auf <https://backend.ebg.admin.ch/fileservice/sdweb-docs-prod-ebgch-files/files/2024/02/21/93a04a37-3772-40c1-a410-8b390ebff5b7.pdf> 2023.
- , "Standard-Analyse-Tool Logib Modul 1, Wegleitung (Version 2023.1)," Eidgenössisches Büro für die Gleichstellung von Frau und Mann EBG. Online verfügbar https://www.ebg.admin.ch/dam/ebg/de/dokumente/lohngleichheit/logib/Wegleitung_Logib_M1.pdf 2023.
- , "Standard-Analyse-Tool Logib Modul 2, Wegleitung (Version 2023.1)," Eidgenössisches Büro für die Gleichstellung von Frau und Mann EBG. Online verfügbar https://www.ebg.admin.ch/dam/ebg/de/dokumente/lohngleichheit/logib/Wegleitung_Logib_M2.pdf 2023.
- Gajta, Patrik**, "Kalibrierung - Beschreibung der "Methode III"," BASS. Interner Entwurf zu Händen des EBG. Mimeo 2023.

A Resultate anderer Umlegungen

Abbildung 18: Vergleich der Resultate M1 und M2

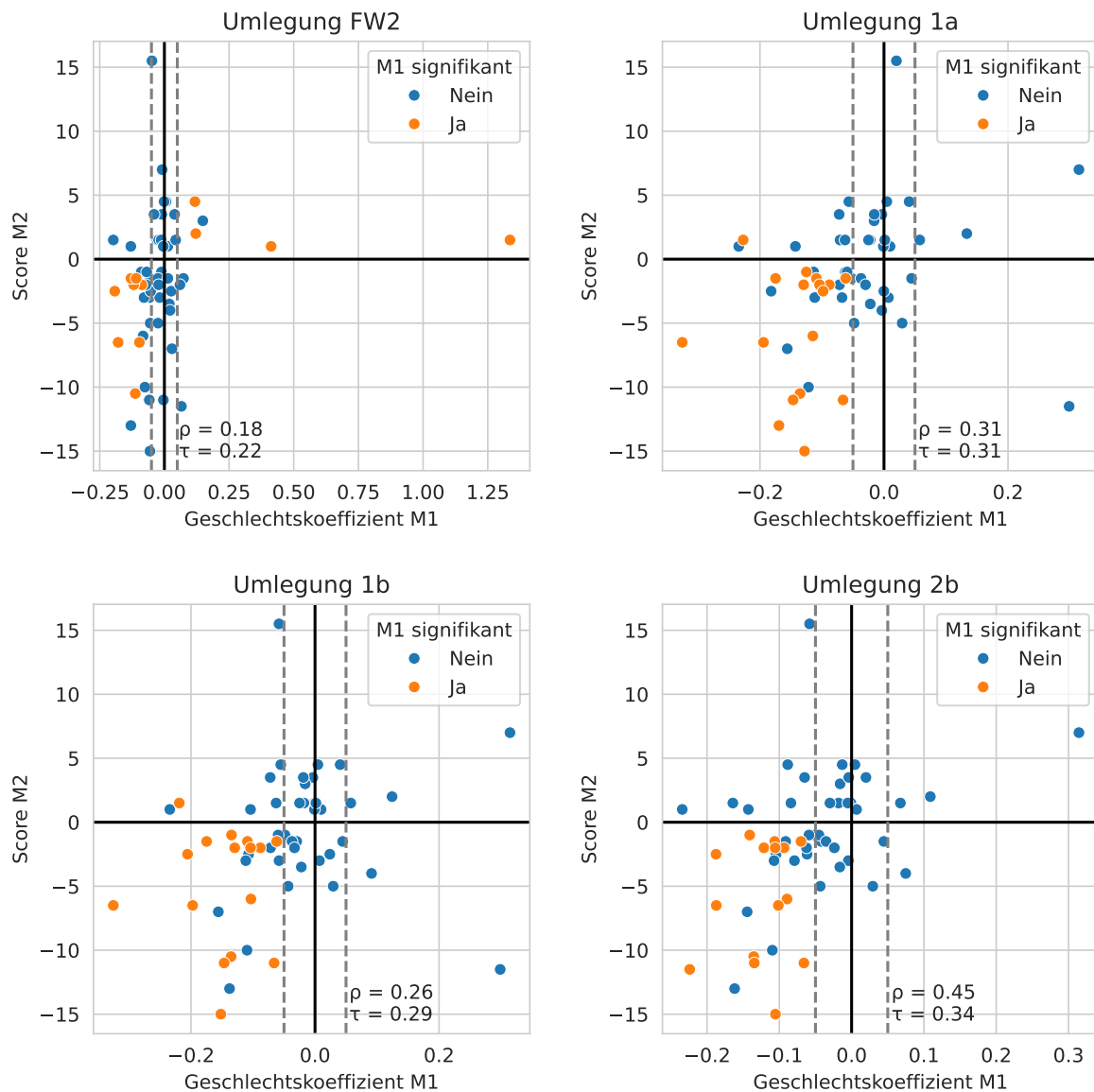


Tabelle 9: Anteil Kontrollen mit Vorzeichenwechsel

Umlegung FW2		
M1 / M2	Zu Ungunsten der Frauen	Zu Ungunsten der Männer
$\beta \leq 0$	49.12%	21.05%
$\beta > 0$	14.04%	15.79%
Umlegung 1a		
M1 / M2	Zu Ungunsten der Frauen	Zu Ungunsten der Männer
$\beta \leq 0$	56.14%	22.81%
$\beta > 0$	7.02%	14.04%
Umlegung 1b		
M1 / M2	Zu Ungunsten der Frauen	Zu Ungunsten der Männer
$\beta \leq 0$	52.63%	24.56%
$\beta > 0$	10.53%	12.28%
Umlegung 2b		
M1 / M2	Zu Ungunsten der Frauen	Zu Ungunsten der Männer
$\beta \leq 0$	57.89%	24.56%
$\beta > 0$	5.26%	12.28%

Tabelle 10: Anzahl Kontrollen mit Vorzeichenwechsel (detailliert)

Umlegung FW2

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	9	4	11	0	0
$-0.05 \leq \beta \leq 0$	2	1	8	3	2
$0 < \beta \leq 0.05$	1	3	3	2	0
$0.05 < \beta$	1	0	5	2	0

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 1a

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	10	4	17	2	0
$-0.05 \leq \beta \leq 0$	1	3	5	3	0
$0 < \beta \leq 0.05$	1	1	3	2	1
$0.05 < \beta$	1	0	2	0	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 1b

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	10	4	14	2	1
$-0.05 \leq \beta \leq 0$	1	1	8	3	0
$0 < \beta \leq 0.05$	1	2	3	2	0
$0.05 < \beta$	1	1	2	0	1

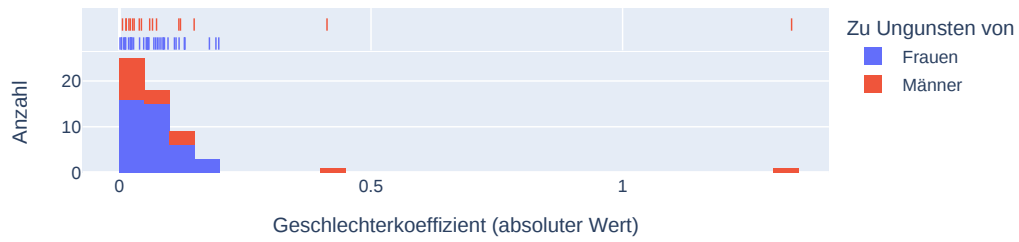
Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 2b

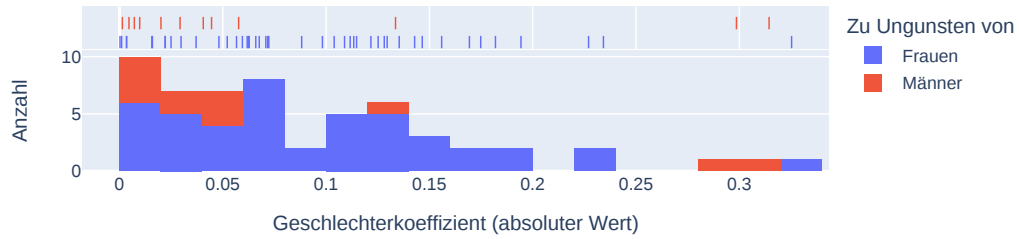
M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	11	5	14	2	1
$-0.05 \leq \beta \leq 0$	1	2	8	3	0
$0 < \beta \leq 0.05$	1	0	3	2	0
$0.05 < \beta$	0	1	2	0	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

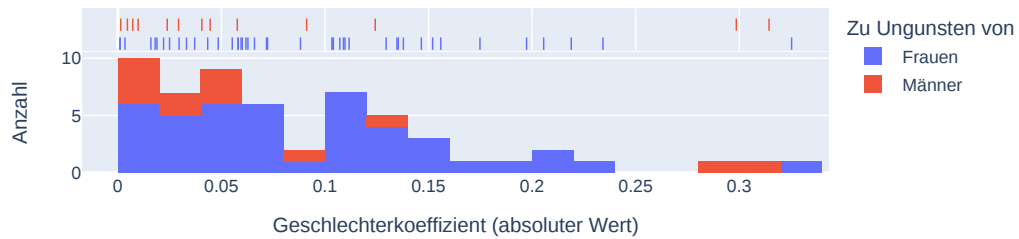
Abbildung 19: Verteilung der Resultate nach Geschlecht
Verteilung beta (Umlegung: FW2)



Verteilung beta (Umlegung: 1a)



Verteilung beta (Umlegung: 1b)



Verteilung beta (Umlegung: 2b)

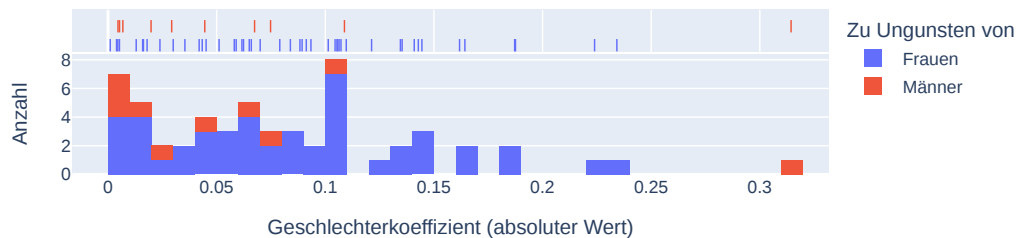


Abbildung 20: Kumulative Verteilung der M2 Kontrollen

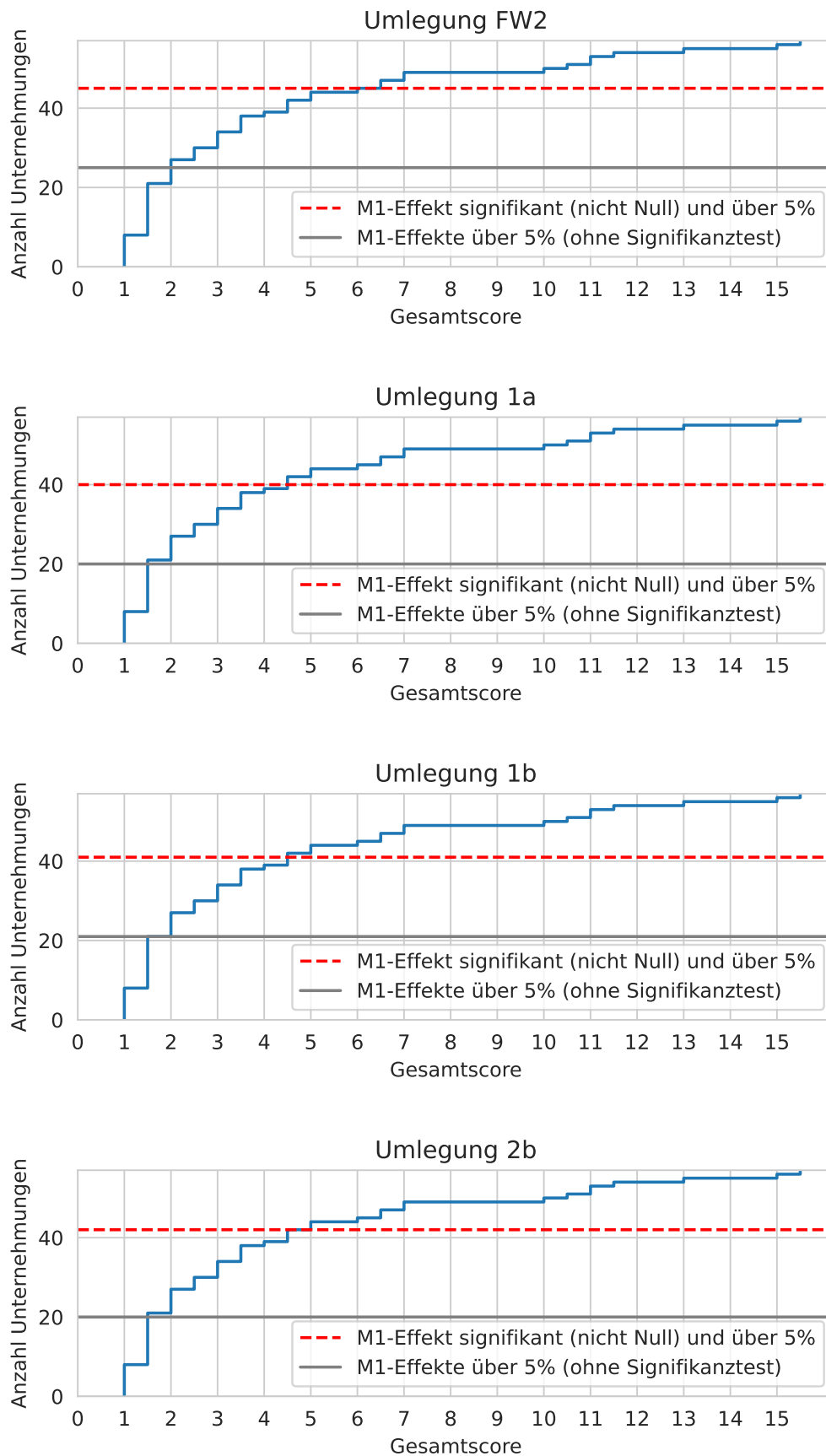
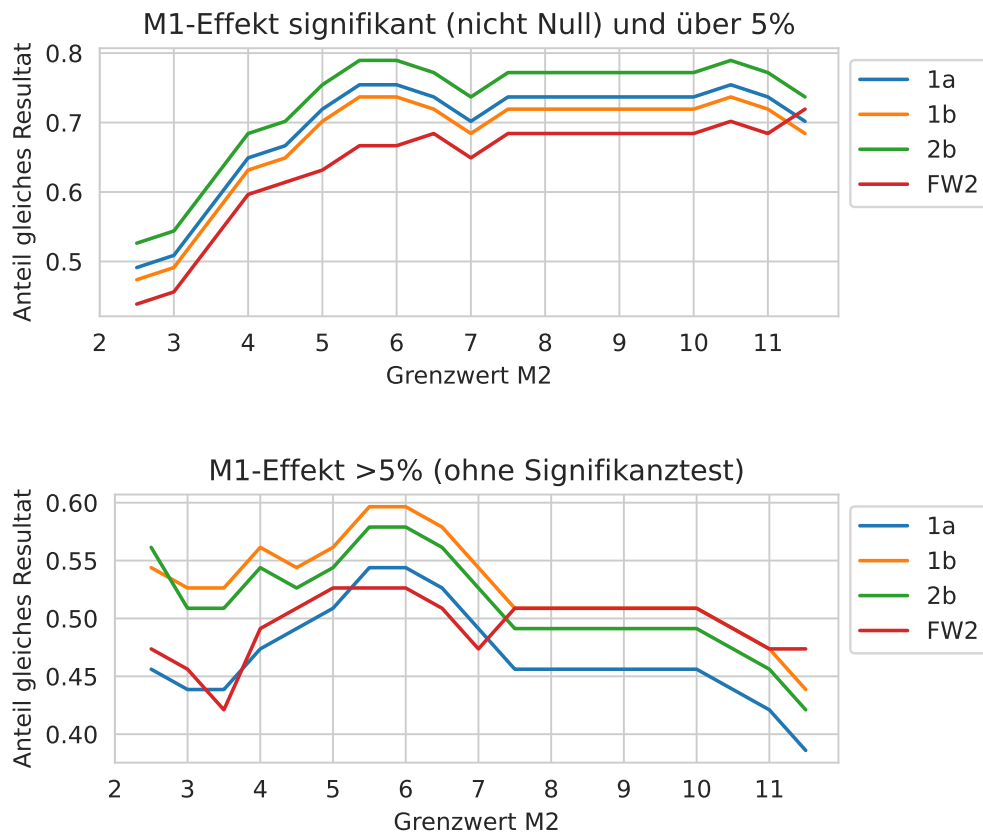


Abbildung 21: Anteil der Kontrollen mit gleichem Resultat M1 und M2



B Resultate anderer Umlegungen (Leave-one-out Ansatz)

Abbildung 22: Vergleich der Resultate M1 und M2 (LOO-Ansatz)

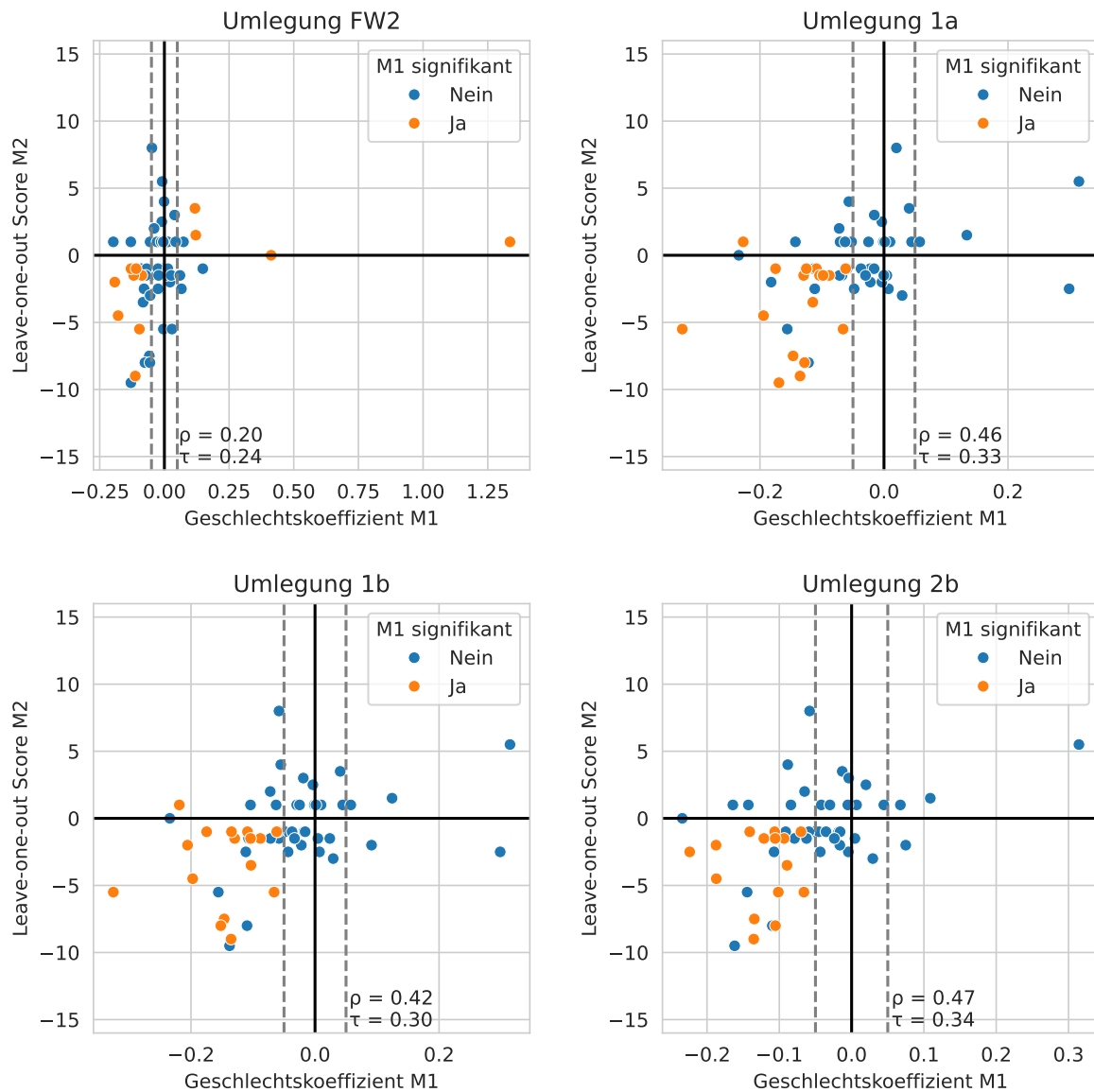


Tabelle 11: Anzahl Kontrollen mit Vorzeichenwechsel (detailliert, LOO-Ansatz))

Umlegung FW2

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	6	4	14	0	0
$-0.05 \leq \beta \leq 0$	1	2	10	1	2
$0 < \beta \leq 0.05$	1	0	7	1	0
$0.05 < \beta$	0	1	6	1	0

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 1a

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	8	3	21	1	0
$-0.05 \leq \beta \leq 0$	0	1	10	1	0
$0 < \beta \leq 0.05$	0	2	4	1	1
$0.05 < \beta$	0	1	2	0	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 1b

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	8	3	18	1	1
$-0.05 \leq \beta \leq 0$	0	1	11	1	0
$0 < \beta \leq 0.05$	0	2	5	1	0
$0.05 < \beta$	0	1	3	0	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Umlegung 2b

M1 / M2	Zu Ungunsten der Frauen		Neutral	Zu Ungunsten der Männer	
	$(-\infty, -5)$	$[-5, -2.5)$		$(2.5, 5]$	$(5, \infty)$
$\beta < -0.05$	8	4	19	1	1
$-0.05 \leq \beta \leq 0$	0	2	10	2	0
$0 < \beta \leq 0.05$	0	1	5	0	0
$0.05 < \beta$	0	0	3	0	1

Es wurden 57 Kontrollen berücksichtigt, somit entspricht 1 Kontrolle 1.75%

Abbildung 23: Kumulative Verteilung der M2 Kontrollen (LOO-Ansatz)

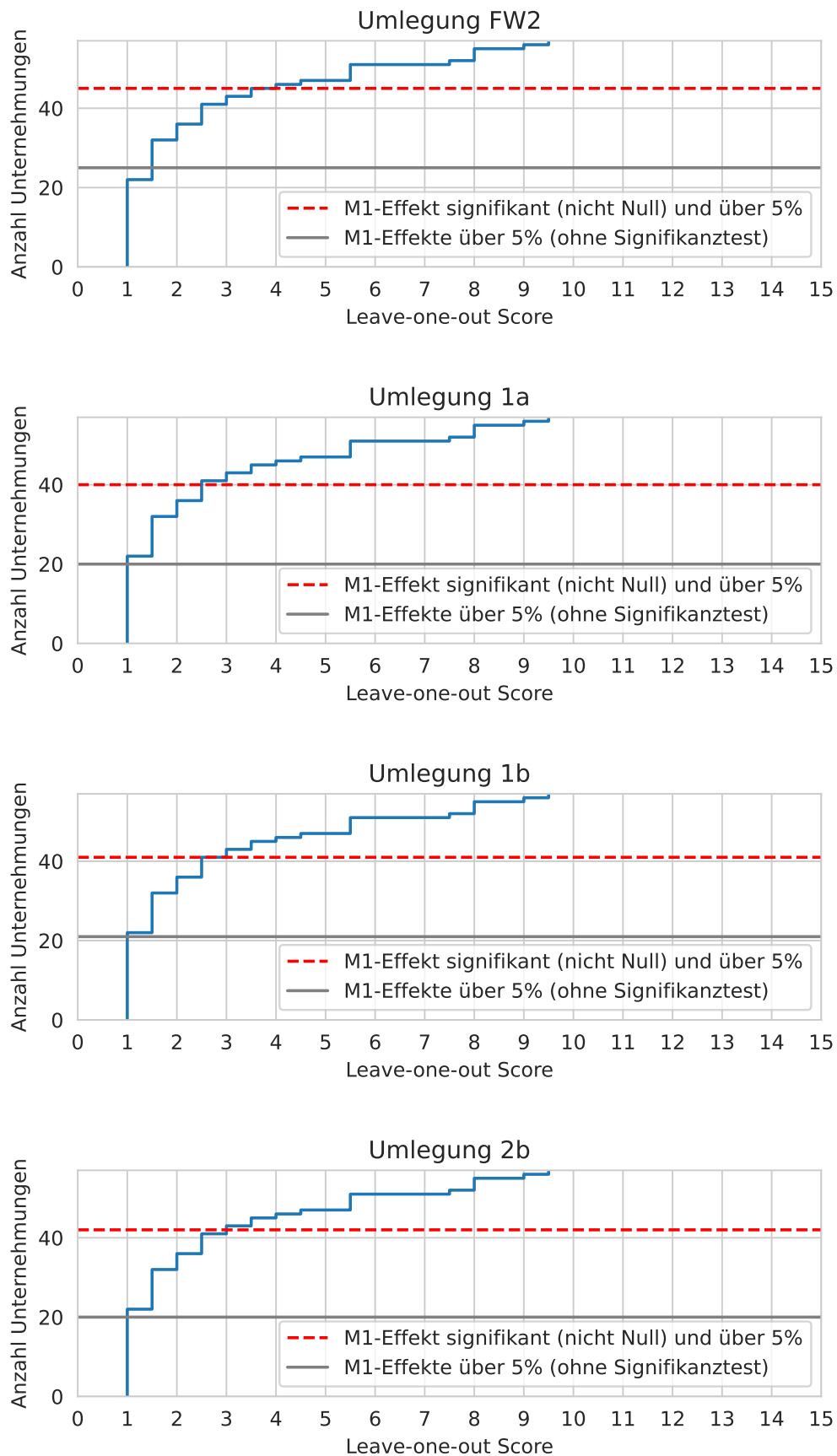


Abbildung 24: Anteil der Kontrollen mit gleichem Resultat M1 und M2 (LOO-Ansatz)

