

Bachelorarbeit im Fach Psychologie

Fehlende Daten in probabilistischen Wissensstrukturen

Eine Simulationsstudie

Julian Mollenhauer

Matrikelnummer: 4093411

29. Oktober 2020

Betreuung: Prof. Dr. Jürgen Heller

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Mathematisch-Naturwissenschaftliche Fakultät

Inhaltsverzeichnis

1	Einleitung	8
2	Probabilistische Wissensstrukturen	9
2.1	BLIM: Parameter und deren Schätzung	10
2.1.1	ML-Schätzung mit EM-Algorithmus	10
3	Umgang mit fehlenden Daten	12
3.1	Ignorable Missingness: IMBLIM	13
3.2	Non-ignorable Missingness: MissBLIM	14
3.3	Sonderfall BLIM	16
4	Parameterschätzung	17
5	Methoden	19
5.1	Simulation der Daten und Wissensstruktur	19
5.1.1	MCAR	20
5.1.2	ks-MNAR	20
5.1.3	iks-MNAR	20
5.2	Empirische Wissensstruktur	21
5.3	Implementierung der Modelle in R	24
5.4	Modellvergleich	24
6	Ergebnisse	25
6.1	Simulierte Wissensstruktur	25
6.1.1	Parameterwiederherstellung	25
6.1.2	Wiederherstellung des latenten Wissenszustandes	29
6.2	Empirische Wissensstrukturen	32
6.2.1	Parameterwiederherstellung	33
6.2.2	Wiederherstellung des latenten Wissenszustandes	40
7	Diskussion	42
A	Herleitung ML-Schätzer	48
B	Items FPI-R	54
C	Abbildungen simulierte Wissensstruktur	56

D	η-Parameter und Stichprobengröße	60
E	Tabellen empirische Wissensstrukturen	61
F	Abbildungen empirische Wissensstrukturen	64

Tabellenverzeichnis

Tabelle 1.	Parameterwerte der Dezimierungsmechanismen für \mathcal{K}_s	21
Tabelle 2.	Fehlende Antworten und Schwelle L der emp. Skalen	22
Tabelle 3.	Parameterwerte der Dezimierungsmechanismen für die emp. Wissensstrukturen.	23
Tabelle 4.	Verzerrung der Parameterschätzung für \mathcal{K}_s	26
Tabelle 5.	Anteil vollständiger Antwortmuster für das BLIM _{COMP}	27
Tabelle 6.	Geschätzte μ - und $\bar{\mu}$ -Parameter des MissBLIM für \mathcal{K}_s	27
Tabelle 7.	Streuung der Parameterschätzung für \mathcal{K}_s	28
Tabelle B1.	FPI-R Items der Skala Gesundheitssorgen	54
Tabelle B2.	FPI-R Items der Skalen Erregbarkeit und Gehemmtheit	55
Tabelle E1.	Verzerrung und Streuung der Parameterschätzungen für \mathcal{K}_{err} ($N = 3700$)	61
Tabelle E2.	Verzerrung und Streuung der Parameterschätzungen für \mathcal{K}_{geh} ($N = 3700$)	62
Tabelle E3.	Verzerrung und Streuung der Parameterschätzungen für \mathcal{K}_{ges} ($N = 3700$)	63

Abbildungsverzeichnis

Abbildung 1.	β -Parameter in Abhängigkeit der Stichprobengröße N für \mathcal{K}_s	30
Abbildung 2.	Mittlere symmetrische Distanzen für \mathcal{K}_s	31
Abbildung 3.	Hasse Diagramm der Precedence-Relation \preceq_{err}	33
Abbildung 4.	Hasse Diagramm der Precedence-Relation \preceq_{geh}	34
Abbildung 5.	Hasse Diagramm der Precedence-Relation \preceq_{ges}	34
Abbildung 6.	β -Parameterschätzung des $BLIM_{COMP}$ für die emp. Wissensstrukturen ($N = 3700$)	36
Abbildung 7.	β -Parameterschätzung des $BLIM_{MAW}$ für die emp. Wissensstrukturen ($N = 3700$)	37
Abbildung 8.	β -Parameterschätzung des $IMBLIM$ für die emp. Wissensstrukturen ($N = 3700$)	38
Abbildung 9.	β -Parameterschätzung des $MissBLIM$ für die emp. Wissensstrukturen ($N = 3700$)	39
Abbildung 10.	Mittlere symmetrische Distanzen für die emp. Wissensstrukturen.	41
Abbildung C1.	Parameterschätzung des $BLIM_{COMP}$ für \mathcal{K}_s	56
Abbildung C2.	Parameterschätzung des $BLIM_{MAW}$ für \mathcal{K}_s	57
Abbildung C3.	Parameterschätzung des $IMBLIM$ für \mathcal{K}_s	58
Abbildung C4.	Parameterschätzung des $MissBLIM$ für \mathcal{K}_s	59
Abbildung D1.	η -Parameter in Abhängigkeit der Stichprobengröße N für \mathcal{K}_s	60
Abbildung F1.	η -Parameterschätzung des $BLIM_{COMP}$ für die emp. Wissensstrukturen ($N = 3700$)	64
Abbildung F2.	η -Parameterschätzung des $BLIM_{MAW}$ für die emp. Wissensstrukturen ($N = 3700$)	65
Abbildung F3.	η -Parameterschätzung des $IMBLIM$ für die emp. Wissensstrukturen ($N = 3700$)	66
Abbildung F4.	η -Parameterschätzung des $MissBLIM$ für die emp. Wissensstrukturen ($N = 3700$)	67
Abbildung F5.	β -Parameterschätzung des $BLIM_{COMP}$ für die emp. Wissensstrukturen ($N = 10000$)	68
Abbildung F6.	β -Parameterschätzung des $BLIM_{MAW}$ für die emp. Wissensstrukturen ($N = 10000$)	69

Abbildung F7. β -Parameterschätzung des IMBLIM für die emp. Wissensstrukturen ($N = 10000$)	70
Abbildung F8. β -Parameterschätzung des MissBLIM für die emp. Wissensstrukturen ($N = 10000$)	71
Abbildung F9. η -Parameterschätzung des BLIM _{COMP} für die emp. Wissensstrukturen ($N = 10000$)	72
Abbildung F10. η -Parameterschätzung des BLIM _{MAW} für die emp. Wissensstrukturen ($N = 10000$)	73
Abbildung F11. η -Parameterschätzung des IMBLIM für die emp. Wissensstrukturen ($N = 10000$)	74
Abbildung F12. η -Parameterschätzung des MissBLIM für die emp. Wissensstrukturen ($N = 10000$)	75
Abbildung F13. Mittlere symmetrische Distanzen für emp. Wissensstrukturen ($N = 10000$)	76

Abstract

Bei der Durchführung psychologischer Tests kommt es häufig dazu, dass Personen einzelne Fragen auslassen und keine Antwort geben. Diese fehlenden Daten stellen bei der Diagnose unterliegender Fähigkeiten ein statistisches Problem dar, da der Prozess, der diese erzeugt oft unbeachtet bleibt. Aufbauend auf eine Arbeit von De Chiusole, Stefanutti, Anselmi und Robusto (2015) wurden zwei Modelle untersucht, die im Rahmen der Theorie der Wissensstrukturen unterschiedliche Ursachen bzw. Prozesse betrachten, die zu fehlenden Daten führen und diese modellieren. Auf Grundlage einer simulierten Wissensstruktur und empirischen Wissensstrukturen aus den Skalen Erregbarkeit, Gehemtheit und Gesundheitssorgen des Freiburger Persönlichkeitsinventars wurden die Modelle hinsichtlich Parameterwiederherstellung und Wiederherstellung des latenten Wissenszustands verglichen. Außerdem wurde ein Vergleich mit dem Basic Local Independence Model mit ausschließlich kompletten Antwortmustern vorgenommen. Es zeigt sich, dass die Modelle, die in der Lage sind die fehlenden Daten zu modellieren besser die Parameter und latenten Wissenszustände wiederherstellen. Dieser Vorteil ist nicht davon abhängig, wie groß die Stichprobe ist, die der Schätzung zugrunde liegt. Weiterhin zeigen sich für das Basic Local Independence Model mit ausschließlich kompletten Antwortmustern starke Verzerrungen bei kleinen Stichproben und großem Anteil fehlender Antworten. Diese verschwinden jedoch mit zunehmender Stichprobengröße bzw. geringerem Anteil fehlender Daten zunehmend. Auch bei Anwendung der sog. Missing-As-Wrong Transformation weist das Basic Local Independence Model systematische Verzerrungen bei der Parameterschätzung und Wiederherstellung des latenten Wissenszustands auf. Alles in allem zeigen sich für die beiden anderen Modelle deutliche Vorteile in der Parameterschätzung und Wiederherstellung des latenten Wissenszustands.

1 Einleitung

Die Theorie der Wissensstrukturen (Doignon & Falmagne, 1985) ist neben der klassischen Testtheorie und der Item-Response Theorie eine weitere Möglichkeit (psychologische) Tests im weitesten Sinne mathematisch zu beschreiben und so deren Eigenschaften und Aussagekraft theoretisch zu erfassen. Sie unterscheidet sich dabei radikal in der Herangehensweise: Es werden nicht aggregierte Daten über die Antworten einer Person hinweg betrachtet, sondern die einzelnen Antwortmuster der Personen bleiben komplett erhalten. Dies führt dazu, dass die Theorie der Wissensstrukturen eine mengentheoretische ist, da sie unter anderem mit den Mengen der richtig gelösten Aufgaben einer Person umgehen muss und diese nicht auf eine einzelne Zahl reduziert. In der vorliegenden Arbeit soll der Umgang mit fehlenden Daten in der Theorie der Wissensstrukturen näher betrachtet werden. Ausgangspunkt bildet die Arbeit von De Chiusole et al. (2015), in welcher entsprechende Modelle hergeleitet wurden sowie deren Simulationsmethodik. Aufbauend darauf werden Eigenschaften der Schätzungen, Verzerrung und Streuung, sowie das Ausgangsmodell näher betrachtet.

Nachfolgend ein kurzer Überblick über die Terminologie: Ausgangspunkt in der Theorie der Wissensstrukturen ist das Paar $\langle Q, \mathcal{K} \rangle$. Q bezeichnet dabei den Wissensbereich, also eine nicht-leere und endliche Menge von zu lösenden Aufgaben $q \in Q$ (oft auch Items genannt) und \mathcal{K} bezeichnet die sogenannte Wissensstruktur. Diese Struktur ist eine Teilmenge der Potenzmenge von Q und beinhaltet alle Wissenszustände K inklusive der Mengen \emptyset und Q . Die Wissenszustände K beschreiben dabei alle möglichen Kombinationen von Aufgaben, die eine Person theoretisch beherrscht und die im Zusammenhang mit diesem Wissensgebiet vorkommen. Die Kardinalität der Wissensstruktur ist oft geringer als die der Potenzmenge von Q , $|\mathcal{K}| \leq 2^{|Q|}$, da aufgrund logischer oder pädagogischer Abhängigkeiten zwischen den Aufgaben gewisse Teilmengen von Q keine Wissenszustände darstellen. Beispielsweise wird eine Person, die in der Lage ist Differenzialgleichungen zu lösen keine Probleme damit haben eine einfache Addition vorzunehmen. Daher ist in einem Wissenszustand K , der die Aufgabe zur Differenzialgleichung enthält auch immer die Aufgabe zur einfachen Addition enthalten und somit wird die Menge der theoretisch möglichen Wissenszustände verkleinert. Diese Abhängigkeiten zwischen den Aufgaben können durch die sogenannte Precedence-Relation \preceq auf Q , mit $a \preceq b$ gdw. „ a eine Voraussetzung“¹

¹Voraussetzung bedeutet hier: Wenn Aufgabe b beherrscht wird, wird auch Aufgabe a be-

ist für b “ für alle $a, b \in Q$, beschrieben werden (Doignon & Falmagne, 1985).

2 Probabilistische Wissensstrukturen

Für eine realistischere Betrachtung des Lösungsverhaltens von Personen muss das Paar $\langle Q, \mathcal{K} \rangle$ erweitert werden. Es wird eine Wahrscheinlichkeitsverteilung $\pi = (\pi_K)_{K \in \mathcal{K}}$ auf den Wissenszuständen $K \in \mathcal{K}$ angenommen. π_K entspricht der Wahrscheinlichkeit, dass eine zufällig ausgewählte Person sich im Wissenszustand K befindet. Außerdem wird die Möglichkeit richtig zu raten und einen Flüchtigkeitsfehler zu begehen mit in die Theorie aufgenommen. Daher muss nun auch das Antwortmuster $R \subseteq Q$, die Menge der tatsächlich gelösten Aufgaben einer Versuchsperson von deren unterliegendem Wissenszustand K unterschieden werden, da nicht mehr alle in K enthaltenen Aufgaben gelöst bzw. nicht in K enthaltenen Aufgaben nicht gelöst werden. Formal lassen sich die Fehler bei einer Aufgabe $q \in Q$ folgendermaßen beschreiben: richtig Raten ($q \in R \wedge q \notin K$) und ein Flüchtigkeitsfehler ($q \notin R \wedge q \in K$). Die Menge aller möglichen Antwortmuster R wird als \mathcal{R} bezeichnet und entspricht im Normalfall der Potenzmenge von Q . Das darauf aufbauende gängigste Modell ist das *Basic Local Independence Model* (BLIM; Doignon & Falmagne, 1999). Die namensgebende Annahme des Modells ist die der lokalen stochastischen Unabhängigkeit: Der Wissenszustand $K \in \mathcal{K}$ liefert eine vollständige Charakterisierung der unterliegenden Fähigkeiten, d.h. bei festem Wissenszustand K ist das Antwortverhalten über alle Aufgaben $q \in Q$ unabhängig. Das Lösen oder nicht-Lösen einer Aufgabe hat somit bei festem Wissenszustand K keinen Einfluss auf das Lösungsverhalten bei einer anderen Aufgabe. Außerdem wird angenommen, dass das Antwortverhalten für jede Aufgabe $q \in Q$ durch einen Parameter β_q (Flüchtigkeitsfehler: Die Person löst die Aufgabe nicht, obwohl sie diese eigentlich beherrscht, $q \notin R \wedge q \in K$) und einen Parameter η_q (Richtig Raten: Person löst die Aufgabe, obwohl sie diese nicht beherrscht, $q \in R \wedge q \notin K$) beschrieben wird. Dabei sind die Parametervektoren $\beta = (\beta_q)_{q \in Q}$ und $\eta = (\eta_q)_{q \in Q}$ unabhängig von den Wissenszuständen in \mathcal{K} , was bedeutet, dass die Wahrscheinlichkeit für einen der beiden Fehler bei einer Aufgabe $q \in Q$ in jedem Wissenszustand K gleich ist. Diese Annahme wird dadurch gerechtfertigt, dass sich ohne sie die Anzahl der Parameter deutlich vergrößern würde (Doignon & Falmagne, 1999).

Ausgehend von $P(R, K)$, der gemeinsamen Wahrscheinlichkeit des Antwortmus-

herrscht, $b \in K \Rightarrow a \in K$

ters R und des Wissenszustands K , lassen sich die Randverteilung π_K und die bedingte Wahrscheinlichkeit $P(R | K)$ betrachten. Dabei besteht für alle $R \in \mathcal{R}$ und $K \in \mathcal{K}$ der Zusammenhang

$$P(R, K) = P(R | K) \cdot \pi_K$$

Aufgrund der obigen Annahmen ergibt sich dann folgende Charakterisierung der bedingten Wahrscheinlichkeit des Antwortmusters R gegeben des latenten Wissenszustands K der Person. Diese wird auch als sog. „Antwortregel“ des BLIM bezeichnet:

$$P(R | K) = \prod_{q \in Q} p_q, \quad p_q = \begin{cases} \beta_q & q \notin R, q \in K \\ 1 - \beta_q & q \in R, q \in K \\ \eta_q & q \in R, q \notin K \\ 1 - \eta_q & q \notin R, q \notin K \end{cases} \quad (1)$$

mit β_q der Wahrscheinlichkeit bei Aufgabe $q \in Q$ einen Flüchtigkeitsfehler zu begehen und η_q der Wahrscheinlichkeit richtig zu Raten.

Für die Wahrscheinlichkeit $P(R)$ eines Antwortmusters $R \in \mathcal{R}$ gilt dann aufgrund des Gesetzes der totalen Wahrscheinlichkeit:

$$P(R) = \sum_{K \in \mathcal{K}} P(R | K) \cdot \pi_K$$

2.1 BLIM: Parameter und deren Schätzung

Die zu schätzenden freien Parameter des BLIM sind einmal die Wahrscheinlichkeiten π_K für alle $K \in \mathcal{K}$ sowie die Fehlerwahrscheinlichkeiten β_q und η_q für alle $q \in Q$. Es ergeben sich somit aufgrund $\sum_{K \in \mathcal{K}} \pi_K \stackrel{!}{=} 1$ insgesamt $|\mathcal{K}| - 1$ freie zu schätzende Wahrscheinlichkeiten $\hat{\pi}_K$ und je $|Q|$ zu schätzende Fehlerwahrscheinlichkeiten $\hat{\beta}_q$ und $\hat{\eta}_q$. Demgegenüber stehen $2^{|Q|} - 1$ unabhängig beobachtbare Häufigkeiten N_R für die Antwortmuster $R \in \mathcal{R}$.

2.1.1 ML-Schätzung mit EM-Algorithmus

Da das BLIM äquivalent zu einem restriktiven Latent-Class Modell ist mit den Wissenszuständen $K \in \mathcal{K}$ als den latenten Klassen (Schrepp, 2005; Ünlü, 2011), kann auch hier der sog. EM-Algorithmus (Dempster, Laird & Rubin, 1977) eingesetzt

werden um die Parameter des Modells zu schätzen (vgl. Stefanutti & Robusto, 2009). Hierbei wird angenommen, dass neben den Antwortmustern R auch noch der latente Wissenszustand K einer Person beobachtet werden kann. Mit diesen Datenpaaren (R, K) , deren absoluten Häufigkeiten $F(R, K)$ und D als der Menge aller Paare (R, K) wird anschließend die Likelihood der vollständigen Daten bestimmt:

$$\mathcal{L}(D \mid \beta, \eta, \pi) = \prod_{R \in \mathcal{R}} \prod_{K \in \mathcal{K}} P(R, K \mid \beta, \eta, \pi)^{F(R, K)}$$

Dann werden die folgenden zwei Schritte iterativ in jedem Durchgang t wiederholt, bis der Algorithmus gegen einen Wert konvergiert und/oder ein bestimmtes Abbruchkriterium erreicht wird:

- **Expectation:** Der Erwartungswert der absoluten (nicht-beobachtbaren) Häufigkeiten $F(R, K)$ wird auf Basis der im vorherigen M-Schritt geschätzten Parameter $\hat{\beta}^{(t-1)}$, $\hat{\eta}^{(t-1)}$ und $\hat{\pi}^{(t-1)}$ berechnet:

$$\mathcal{E}(F(R, K)) = N_R \cdot P(K \mid R, \hat{\beta}^{(t-1)}, \hat{\eta}^{(t-1)}, \hat{\pi}^{(t-1)})$$

- **Maximization:** Mit $F(R, K) = \mathcal{E}(F(R, K))$ werden die neuen Maximum-Likelihood Schätzer $\hat{\beta}^{(t)}$, $\hat{\eta}^{(t)}$ und $\hat{\pi}^{(t)}$ berechnet. Für die ML-Schätzungen in Durchgang t ergeben sich folgende Gleichungen (Herleitung: siehe Anhang A):

$$\begin{aligned} \hat{\beta}_q^{(t)} &= \frac{\sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}} F(R, K) \cdot i_{q \in (K \setminus R)}}{\sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}} F(R, K) \cdot i_{q \in K}} \\ \hat{\eta}_q^{(t)} &= \frac{\sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}} F(R, K) \cdot i_{q \in (R \setminus K)}}{\sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}} F(R, K) \cdot i_{q \in (Q \setminus K)}} \\ \hat{\pi}_K^{(t)} &= \frac{1}{N} \sum_{R \in \mathcal{R}} F(R, K) \end{aligned}$$

mit der Stichprobengröße $N = \sum_{R \in \mathcal{R}} \sum_{K \in \mathcal{K}} F(R, K)$ und den Indikatorvariablen i_{\dots} , die nach dem Schema $i_{q \in \mathcal{T}}$ mit $\mathcal{T} = K \setminus R$, $\mathcal{T} = K$, $\mathcal{T} = R \setminus K$ und $\mathcal{T} = Q \setminus K$ aufgebaut sind:

$$i_{q \in \mathcal{T}} = \begin{cases} 1 & , \text{wenn } q \in \mathcal{T} \\ 0 & \text{sonst} \end{cases} \quad (2)$$

3 Umgang mit fehlenden Daten

Bei der praktischen Anwendung von Tests, Fragebögen u.ä. ist es häufig der Fall, dass Personen nicht alle Aufgaben bearbeitet haben, also weder eine richtige noch falsche Antwort vorliegt. Gründe hierfür sind vielfältig und reichen von „eine Person gibt lieber keine, statt eine falsche Antwort, also nicht-Wissen“ bis „aufgrund von Zeitdruck können nicht alle Aufgaben bearbeitet werden“. Diese fehlenden Daten müssen bei der Auswertung berücksichtigt werden, damit es zu keinen systematischen Verzerrungen bei der Parameterschätzungen und Wiederherstellung der latenten Wissenszustände kommt. Dafür ist wesentlich, wie diese fehlenden Daten zustande kommen, also ob es z.B. Abhängigkeiten zwischen den fehlenden Daten und richtig gelösten Aufgaben gibt. Die verschiedenen Prozesse, die zu fehlenden Daten führen, werden im folgenden auch als Dezimierungsmechanismen bezeichnet. Little und Rubin (2002) unterscheiden drei wesentliche statistische Mechanismen, die zu fehlenden Daten führen. Für deren Verständnis betrachten wir eine $n \times m$ Matrix Y , deren Spalten die verschiedenen Aufgaben und Zeilen die verschiedenen Personen darstellen. Die Antwort einer Person i auf die Aufgabe j wird in Y mit $y_{ij} = 1$ kodiert, wenn Person i Aufgabe j löst, sonst mit 0. Diese theoretische Matrix Y enthält die kompletten Daten. Zusätzlich gibt es eine $n \times m$ Matrix M , in der eine fehlende Antwort von Person i auf Aufgabe j mit $m_{ij} = 1$ und sonst 0 kodiert wird. Die Indikatorvariablen in M für fehlende Daten werden dann als Zufallsvariablen betrachtet um anhand der bedingten Wahrscheinlichkeit $P(M | Y)$ den Zusammenhang zwischen M und Y zu charakterisieren. Dieser ist ausschlaggebend für die Unterscheidung der verschiedenen Dezimierungsmechanismen:

- Missing Completely at Random (MCAR): In diesem Fall gibt es keinen Zusammenhang zwischen den fehlenden Daten M und den vollständigen Daten Y , somit gilt $P(M | Y) = P(M)$. Dieser Fall wird auch als *ignorable missings* bezeichnet. Angewandt auf die Terminologie der Theorie der Wissensstrukturen gilt bei MCAR, dass der Dezimierungsmechanismus unabhängig vom Wissenszustand K ist. Dies ist z.B. dann der Fall, wenn Personen nur eine zufällige Teilmenge von Aufgaben aus Q zur Bearbeitung vorgelegt wird.
- Missing at Random (MAR): Hier wird die Matrix Y in einen beobachtbaren Teil Y_{obs} und nicht-beobachtbaren Teil Y_{miss} aufgeteilt. Letzterer kennzeichnet das theoretische Antwortverhalten bei den nicht beantworteten Aufgaben. Bei MAR hängt M nur von den beobachteten Daten Y_{obs} ab, es gilt also $P(M |$

$Y) = P(M | Y_{obs})$. Dieser Fall wird hier jedoch nicht weiter behandelt werden.

- Not Missing at Random (MNAR²): Hier wird angenommen, dass die fehlenden Daten M nicht unabhängig von Y_{miss} sind. Man spricht daher auch von *non-ignorable missings*. Bei MNAR werden nun Abhängigkeiten zwischen den fehlenden Antworten M und dem Wissenszustand K angenommen. Dies bedeutet, dass abhängig von den beherrschten Aufgaben in K , die fehlenden Daten M variieren. Wenn man beispielsweise annimmt, dass Personen, die eine Aufgabe q nicht beherrschen, $q \notin K$, eher dazu neigen keine Antwort zu geben, als Personen, die q beherrschen.

De Chiusole et al. (2015) haben für die beiden Fälle MCAR und MNAR Abwandlungen des BLIM entwickelt, die die Modellierung der fehlenden Daten unter MCAR und MNAR Bedingung ermöglichen. Hierzu muss neben dem Antwortmuster R , der Menge der richtig beantworteten Aufgaben $q \in Q$, auch die Menge der fehlenden Antworten $M \subseteq (Q \setminus R)$ betrachtet werden. Alle möglichen Paare $\langle R, M \rangle$ werden in der Menge \mathcal{Q} zusammengefasst und stellen, neben den absoluten Häufigkeiten $F(R, M)$ die beobachtbaren Daten dar, $\mathcal{Q} = \{\langle R, M \rangle : R \subseteq Q, M \subseteq (Q \setminus R)\}$. Für die Likelihood dieser Daten gilt dann:

$$\mathcal{L}(\mathcal{Q} | \beta, \eta, \pi) = \prod_{\langle R, M \rangle \in \mathcal{Q}} P(R, M | \beta, \eta, \pi)^{F(R, M)}$$

Die nachfolgend beschriebenen Modelle unterscheiden sich dann in der Berechnung der Wahrscheinlichkeiten $P(R, M)$.

3.1 Ignorable Missingness: IMBLIM

Das *Ignorable Missingness BLIM* (IMBLIM) beruht auf der Annahme, dass die fehlenden Antworten der Personen dem MCAR Dezimierungsmechanismus entsprechen. Das heißt, dass hier keine Abhängigkeiten zwischen den fehlenden Antworten M und dem Wissenszustand K angenommen werden. Unabhängig in welchem Wissenszustand eine Person sich befindet und somit auch unabhängig von der einzelnen Aufgabe fehlen die Daten mit einer gewissen Wahrscheinlichkeit. Der Dezimierungsmechanismus wird also als unabhängig von dem Prozess angesehen, der die theoretisch vollständigen Daten R^* erzeugt (De Chiusole et al., 2015). Für die

²Little und Rubin (2002) haben die Abkürzung NMAR gewählt, da aber in dem Paper von De Chiusole et al. (2015), auf das diese Arbeit größtenteils aufbaut, die Bezeichnung MNAR verwendet wird, soll diese hier auch so gewählt werden.

Modellierung der fehlenden Antworten wird nun ein zweistufiger Prozess angenommen: Im ersten Schritt werden basierend auf dem BLIM und allen damit einhergehenden Annahmen (siehe Abschnitt 2) die vollständigen Daten R^* modelliert. Der zweite Schritt besteht in der Anwendung des Dezimierungsmechanismus, der aus Q die fehlenden Daten M „auswählt“. Es gilt dann: $R = R^* \setminus M$ und man erhält das beobachtbare Paar $\langle R, M \rangle$. Mathematisch bedeutet dies, dass zuerst die Wahrscheinlichkeiten $P(R^*, M | K)$ betrachtet werden können, also die Wahrscheinlichkeiten der vollständigen Daten R^* und der fehlenden Daten M gegeben die Person befindet sich im Wissenszustand K . Aufgrund der Annahme, dass der Dezimierungsmechanismus unabhängig von dem Prozess ist, der die vollständigen Daten R^* erzeugt, gilt: $P(R^*, M | K) = P(R^* | K) \cdot P(M)$. Unter der plausiblen Annahme, dass $P(M)$ unabhängig von den Parametern β und η ist, folgt daraus dann $P(R^*, M) = P(M) \sum_{K \in \mathcal{K}} P(R^* | K) \cdot \pi_K$. De Chiusole et al. (2015) haben gezeigt, dass darauf aufbauend für die Wahrscheinlichkeit der Daten $\langle R, M \rangle$ gilt:

$$P(R, M) = P(M) \sum_{K \in \mathcal{K}} P(\langle R, M \rangle^* | K) \cdot \pi_K \quad (3)$$

mit:

$$P(\langle R, M \rangle^* | K) = \prod_{q \in Q} p_q, \quad p_q = \begin{cases} \beta_q & q \notin R, q \in K, q \notin M \\ 1 - \beta_q & q \in R, q \in K, q \notin M \\ \eta_q & q \in R, q \notin K, q \notin M \\ 1 - \eta_q & q \notin R, q \notin K, q \notin M \end{cases} \quad (4)$$

wobei $\langle R, M \rangle^*$ die Menge der zum Paar $\langle R, M \rangle$ kompatiblen Antwortmuster darstellt, d.h. $\langle R, M \rangle^* = \{R \cup C : C \subseteq M\}$. Es wird hier also strukturell die Antwortregel des BLIM (vgl. Gleichung (1)) angewendet, nur dass die Aufgaben mit fehlenden Antworten nicht berücksichtigt werden.

Bei keinen fehlenden Daten, also $M = \emptyset$, sind die Vorhersagen für IMBLIM und BLIM identisch, da wie oben leicht zu sehen, die Wahrscheinlichkeit $P(\langle R, M \rangle^* | K)$ zu $P(R | K)$ wird.

3.2 Non-ignorable Missingness: MissBLIM

De Chiusole et al. (2015) haben auch für den MNAR Dezimierungsmechanismus eine Erweiterung des BLIM hergeleitet. Hierbei wird nun die Unabhängigkeit zwischen den fehlenden Antworten M und dem Wissenszustand K aufgegeben. Je nach Wissenszustand K ist die Wahrscheinlichkeit, dass ein bestimmtes Item $q \in Q$ nicht

beantwortet wird verschieden. Dies ermöglicht eine viel flexiblere Modellierung der fehlenden Antworten, hat aber auch einen Anstieg der Parameter zur Folge. Insgesamt werden zusätzlich $2 \cdot |Q|$ Parameter benötigt. Es lässt sich aufgrund der Nähe zu Latent-Class-Modellen folgendes Modell aufstellen (De Chiusole et al., 2015):

$$P(R, M) = \sum_{K \in \mathcal{K}} P(\langle R, M \rangle^* | K) \cdot P(M | K) \cdot \pi_K \quad (5)$$

mit $P(\langle R, M \rangle^* | K)$ wie in Gleichung (4) und der bedingten Wahrscheinlichkeit der fehlenden Antworten M gegeben den latenten Wissenszustand K , $P(M | K)$. Die Abhängigkeit zwischen M und K führt dazu, dass die Wahrscheinlichkeit $P(M | K)$ definiert werden muss. Hierzu sind zwei Annahmen zu treffen: Erstens ist wie schon bei dem Antwortmuster R bei der Anwendung des BLIM auch für die fehlenden Antworten M die Annahme der lokalen stochastischen Unabhängigkeit wichtig. Sie besagt, dass bei festem Wissenszustand K das nicht-Beantworten bzw. Beantworten einer Aufgabe keinen Einfluss auf das Antwortverhalten bei den anderen Aufgaben hat. Die gesamte Information über die fehlenden Daten steckt somit in dem Wissenszustand K . Die zweite Annahme beschreibt die mathematische Modellierung und ist eng mit der ersten verknüpft. Es wird angenommen, dass $P(M | K)$ folgendermaßen beschrieben werden kann (De Chiusole et al., 2015):

$$P(M | K) = \prod_{q \in Q} p_q, \quad p_q = \begin{cases} \mu_q & q \in M, q \in K \\ 1 - \mu_q & q \notin M, q \in K \\ \mu_{\bar{q}} & q \in M, q \notin K \\ 1 - \mu_{\bar{q}} & q \notin M, q \notin K \end{cases} \quad (6)$$

μ_q beschreibt hier die Wahrscheinlichkeit, dass Aufgabe q , wenn sie in K enthalten ist, nicht beantwortet wird, und $\mu_{\bar{q}}$ beschreibt die Wahrscheinlichkeit, dass die Aufgabe nicht beantwortet wird, wenn sie nicht in K enthalten ist. Über diese Parameter $\mu = (\mu_q)_{q \in Q}$ und $\bar{\mu} = (\mu_{\bar{q}})_{q \in Q}$ lässt sich somit die Abhängigkeit zwischen M und K erfassen.

Für den Zusammenhang zwischen BLIM und MissBLIM gilt, dass sie äquivalent sind, sobald keine fehlenden Antworten existieren, also $M = \emptyset$ gilt. Dies lässt sich für $P(\langle R, M \rangle^* | K)$ leicht anhand der Beziehung zwischen BLIM und IMBLIM zeigen. Weiterhin gilt für $P(M | K)$, dass aufgrund $M = \emptyset$ die Parameter μ_q und $\mu_{\bar{q}}$ gleich 0 sind und somit nur die Paare $\langle R, M \rangle$, bei denen $M = \emptyset$ eine Wahrscheinlichkeit $P(R, M) > 0$ haben (De Chiusole et al., 2015). Bei vollständigen Daten sind also

alle drei Modelle, BLIM, IMBLIM und MissBLIM, äquivalent.

3.3 Sonderfall BLIM

Da das BLIM in der ursprünglichen Form nicht mit fehlenden Daten umgehen kann, müssen die fehlenden Antworten transformiert oder für die Schätzung der Parameter des BLIM ausgeschlossen werden. Das Ausschließen der Antwortmuster, die fehlende Daten enthalten, kann dazu führen, dass die Schätzungen sehr ungenau werden, da die Stichprobengröße mit zunehmendem Anteil fehlender Daten abnimmt. Auf der anderen Seite ist eine Transformation der fehlenden Werte immer mit Annahmen verbunden, die, wenn sie nicht gerechtfertigt sind, zu einer Verzerrung der Schätzungen führen können. Die gängigste Annahme hierbei ist, dass fehlende Daten aufgrund nicht-Wissen zustande kommen. Eine Person beantwortet danach eine Frage nicht, wenn sie diese nicht beherrscht. Die anzuwendende Transformation ist dann die „Missing-as-Wrong“-Transformation (im Folgenden: MAW-Transformation). Fehlende Antworten werden also der Menge der nicht gelösten Aufgaben zugeordnet. De Chiusole et al. (2015) haben gezeigt, dass die MAW-Transformation bei Daten, die dem im MissBLIM postulierten Prozess entstammen, zu systematischen Verzerrungen der Schätzungen $\hat{\beta}_q$ und $\hat{\eta}_q$ führt. Diese Verzerrungen sind proportional zu den μ_q und $\mu_{\bar{q}}$ Parametern. Es gilt mit β_q^{true} und η_q^{true} als den wahren Werten der Parameter β_q und η_q :

$$\Delta\beta_q^{MAW} = \hat{\beta}_q - \beta_q^{true} = \mu_q \cdot (1 - \beta_q^{true}) \geq 0 \quad (7)$$

$$\Delta\eta_q^{MAW} = \hat{\eta}_q - \eta_q^{true} = -\mu_{\bar{q}} \cdot \eta_q^{true} \leq 0 \quad (8)$$

Die beiden Spezialfälle des BLIM sowohl mit MAW-Transformation als auch mit ausschließlich den vollständigen Antwortmustern werden im folgenden als $BLIM_{MAW}$ bzw. $BLIM_{COMP}$ bezeichnet.

4 Parameterschätzung

Wie schon für das BLIM (vgl. Abschnitt 2.1.1) kann die Schätzung der Parameter des IMBLIM sowie des MissBLIM mithilfe des EM-Algorithmus erfolgen (De Chiusole et al., 2015). Es wird hierbei angenommen, dass neben dem Paar $\langle R, M \rangle$ auch der jeweils zugehörige Wissenszustand K beobachtbar ist. Für die Schätzung muss die Likelihood der vollständigen Daten aufgestellt werden. Dann können die erwarteten absoluten Häufigkeiten $F(R, M, K)$ im Durchgang t mit den Parameterschätzungen des vorherigen Durchgangs $t - 1$ berechnet und die neuen Maximum-Likelihood-Schätzer geschätzt werden. Die letzten beiden Schritte werden nun iterativ so lange wiederholt, bis der Algorithmus gegen einen Wert konvergiert und/oder ein bestimmtes Abbruchkriterium erreicht wird.

Für die Likelihood der vollständigen Daten \mathcal{Q}_K , als der Menge aller mögliche Tripel $\langle R, M, K \rangle$ und dem Parametervektor θ , der alle Parameter des Modells enthält, gilt allgemein:

$$\mathcal{L}(\mathcal{Q}_K | \theta) = \prod_{\langle R, M, K \rangle \in \mathcal{Q}_K} P(R, M, K | \theta)^{F(R, M, K)}$$

Für das IMBLIM folgt daraus mit den Parametervektoren β , η und π für alle Aufgaben $q \in Q$ bzw. Wissenszustände $K \in \mathcal{K}$

$$\mathcal{L}(\mathcal{Q}_K | \beta, \eta, \pi) = \prod_{\langle R, M, K \rangle \in \mathcal{Q}_K} (P(M) \cdot P(\langle R, M \rangle^* | K, \beta, \eta) \cdot \pi_K)^{F(R, M, K)} \quad (9)$$

und für das MissBLIM gilt mit zusätzlich den Parametervektoren μ für alle Parameter μ_q und $\bar{\mu}$ für alle Parameter $\mu_{\bar{q}}$

$$\mathcal{L}(\mathcal{Q}_K | \beta, \eta, \pi, \mu, \bar{\mu}) = \prod_{\langle R, M, K \rangle \in \mathcal{Q}_K} (P(\langle R, M \rangle^* | K, \beta, \eta) \cdot P(M | K, \mu, \bar{\mu}) \cdot \pi_K)^{F(R, M, K)} \quad (10)$$

Der benötigte Erwartungswert für die absoluten nicht-beobachtbaren Häufigkeiten $F(R, M, K)$ im Durchgang t mit den Parameterschätzungen $\hat{\theta}^{(t-1)}$ aus dem vorherigen Durchgang und den beobachtbaren absoluten Häufigkeiten $F(R, M)$ der richtig gelösten Aufgaben R und fehlenden Antworten M lässt sich allgemein folgendermaßen berechnen:

$$\mathcal{E}(F(R, M, K)) = F(R, M) \cdot P(K | R, M, \hat{\theta}^{(t-1)}) \quad (11)$$

Die Wahrscheinlichkeit $P(K | R, M, \hat{\theta}^{(t-1)})$ lässt sich dann für das IMBLIM und MissBLIM mit Hilfe des Satz von Bayes berechnen

$$P(K | R, M, \hat{\theta}^{(t-1)}) = \frac{P(R, M | K, \hat{\theta}^{(t-1)}) \cdot \pi_K}{P(R, M | \hat{\theta}^{(t-1)})} \quad (12)$$

mit $P(R, M | K) = P(\langle R, M \rangle^* | K, \hat{\beta}^{(t-1)}, \hat{\eta}^{(t-1)}) \cdot P(M)$ für das IMBLIM mit $P(\langle R, M \rangle^* | K, \hat{\beta}^{(t-1)}, \hat{\eta}^{(t-1)})$ wie in Gleichung (4) und $P(M)$ als der relativen Häufigkeit der fehlenden Antworten M . Für das MissBLIM gilt entsprechend $P(R, M | K) = P(\langle R, M \rangle^* | K, \beta^{(t-1)}, \eta^{(t-1)}) \cdot P(M | K, \mu^{(t-1)}, \bar{\mu}^{(t-1)})$ mit den Gleichungen (4) und (6). Weiterhin ist $P(R, M)$ in Gleichung (3) für das IMBLIM bzw. in Gleichung (5) für das MissBLIM definiert.

Mit diesen erwarteten nicht-beobachtbaren Häufigkeiten $\mathcal{E}(F(R, M, K))$ für $F(R, M, K)$ in die Likelihood (9) bzw. (10) eingesetzt, können für den Durchgang t nun neue ML-Parameterschätzungen berechnet werden. Hierfür ist die Likelihood partiell nach den einzelnen Parametern abzuleiten, Null zu setzen und anschließend nach den Parametern aufzulösen. Es ergeben sich für das IMBLIM folgende Schätzungen für die Parameter im Durchgang t , wobei die Indikatorvariablen i_{\dots} nach dem gleichen Schema wie in Gleichung (2) aufgebaut sind (für eine Herleitung der ML-Schätzer siehe Anhang A):

$$\hat{\beta}_q^{(t)} = \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \setminus (R \cup M)}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \setminus M}} \quad (13)$$

$$\hat{\eta}_q^{(t)} = \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in R \setminus K}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in Q \setminus (M \cup K)}} \quad (14)$$

$$\hat{\pi}_K^{(t)} = \frac{1}{N} \sum_{\langle R, M \rangle \in \mathcal{Q}} F(R, M, K) \quad (15)$$

mit der Stichprobengröße $N = \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K)$.

Für das MissBLIM ergeben sich für β , η und π die gleichen ML-Schätzer wie in den Gleichungen (13), (14) und (15) für das IMBLIM beschrieben. Hinzu kommen jedoch noch die Schätzungen für die Parameter μ und $\bar{\mu}$:

$$\hat{\mu}_q^{(t)} = \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \cap M}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K}} \quad (16)$$

$$\hat{\mu}_{\bar{q}}^{(t)} = \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in M \setminus K}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in Q \setminus K}} \quad (17)$$

5 Methoden

Die Anpassungen des BLIM für den Umgang mit fehlenden Daten wurden zuerst wie bei De Chiusole et al. (2015) anhand einer simulierten Wissensstruktur verglichen. Um jedoch zu überprüfen, ob diese Befunde sich auch basierend auf einer empirischen Wissensstruktur zeigen, wurden darüber hinaus drei empirische Wissensstrukturen aus der Normstichprobe des Freiburger Persönlichkeitsinventars (Fahrenberg, Hampel & Selg, 2010) hergeleitet.

5.1 Simulation der Daten und Wissensstruktur

Für den Vergleich der Modelle wurden verschiedene Datensätze simuliert. Dabei wurden äquivalent zu dem Vorgehen von De Chiusole et al. (2015) drei verschiedene Bedingungen betrachtet, wie die fehlenden Antworten zustande kommen können. Ausgangspunkt bildete eine simulierte Wissensstruktur \mathcal{K}_s auf der Menge der Aufgaben Q und eine Wahrscheinlichkeitsverteilung π auf \mathcal{K}_s . Diese Wissensstruktur wurde zufällig aus der Potenzmenge $\mathcal{P}(Q)$ mit $|Q| = 25$ erzeugt, indem ohne zurücklegen zufällig 500 Elemente aus $\mathcal{P}(Q)$ gezogen wurden. Die Wahrscheinlichkeitsverteilung π wurde generiert indem 500 mal aus der Gleichverteilung $\text{unif}(0, 1)$ gezogen wurde und diese Werte dann auf 1 summennormiert wurden. Die Parameter β und η wurden jeweils aus der Gleichverteilung $\text{unif}(0, 0.1)$ gezogen. Zwischen den verschiedenen simulierten Datensätzen variierten die Parameterwerte für die β -, η - und π -Parameter nicht.

Die Simulation der Daten einer Person erfolgt folgendermaßen:

1. Ein Wissenszustand $K \in \mathcal{K}_s$ wird mit Wahrscheinlichkeit π_K zufällig gezogen.
2. Basierend auf diesem Wissenszustand K wird mit Hilfe des BLIM und den zugehörigen Parametern β und η das vollständige Antwortmuster R^* simuliert.
3. Ein Dezimierungsmechanismus wählt aus Q die Menge der fehlenden Antworten M . Dafür werden die Wahrscheinlichkeiten $P(q \in M \mid q \in K) = \mu_q$ und $P(q \in M \mid q \notin K) = \mu_{\bar{q}}$ betrachtet. Diese korrespondieren zu den entsprechenden Parametern des MissBLIM und geben an, mit welcher Wahrscheinlichkeit eine Aufgabe, die in K enthalten ist, nicht beantwortet wird, $P(q \in M \mid q \in K)$, bzw. mit welcher Wahrscheinlichkeit eine Aufgabe, die nicht in K enthalten ist, nicht beantwortet wird, $P(q \in M \mid q \notin K)$.
4. Mit $R = R^* \setminus M$ erhält man dann das beobachtbare Antwortmuster R .

Für jede simulierte Person ergibt sich somit ein Tripel bestehend aus Antwortmuster R , fehlenden Antworten M und latentem Wissenszustand K . Ein simulierter Datensatz entsteht dann durch das N -fache Ausführen der Schritte 1-4. N ist dementsprechend die Stichprobengröße. In Schritt 3 wurden verschiedene Dezimierungsmechanismen verwendet, die im folgenden erläutert und in Tabelle 1 zusammengefasst sind.

5.1.1 MCAR

Für die MCAR Bedingung (im Folgenden auch nur mc genannt) wird angenommen, dass der Dezimierungsmechanismus unabhängig vom Wissenszustand K ist. Die Menge der fehlenden Antworten M entsteht dann dadurch, dass jede Aufgabe $q \in Q$ mit einer festen Wahrscheinlichkeit in die Menge M aufgenommen wird. In einem Datensatz war diese Wahrscheinlichkeit immer gleich, nahm aber über die Datensätze hinweg folgende Werte an: $\{0.1, 0.2, 0.3, 0.4\}$.

5.1.2 ks-MNAR

In der ks-MNAR Bedingung (im Folgenden auch nur ks genannt) wird der Fall simuliert, dass Personen ein Item, welches sie nicht beherrschen ($q \notin K$), mit einer gewissen Wahrscheinlichkeit nicht beantworten. Dafür wurde jede Aufgabe $q \notin K$ mit einer Wahrscheinlichkeit der Menge M zugeordnet. Über die Datensätze hinweg nahm diese Wahrscheinlichkeit folgende Werte an: $\{0.2, 0.4, 0.6, 0.8\}$. Diese Wahrscheinlichkeiten entsprechen dem doppelten derer bei MCAR um den gleichen Anteil an fehlenden Antworten zu erzeugen.

5.1.3 iks-MNAR

In der iks-MNAR Bedingung (im Folgenden auch nur iks genannt) variiert nun sowohl die Wahrscheinlichkeit, dass eine Aufgabe, die nicht im Wissenszustand K enthalten ist, nicht beantwortet wird, als auch die Wahrscheinlichkeit, dass eine Aufgabe, die im Wissenszustand K enthalten ist, keine Antwort erhält. Dafür wurden die Parameter $P(q \in M \mid q \in K) = \mu_q$ und $P(q \in M \mid q \notin K) = \mu_{\bar{q}}$ aus Gleichverteilungen gezogen. Insgesamt wurden fünf Kombinationen $C1$ bis $C5$ betrachtet. Die genauen Gleichverteilungen für μ_q und $\mu_{\bar{q}}$ sind Tabelle 1 zu entnehmen.

Tabelle 1

Parameterwerte μ_q und $\mu_{\bar{q}}$ der verschiedenen Dezimierungsmechanismen für die Datensätze basierend auf der simulierten Wissensstruktur \mathcal{K}_s .

	MCAR		ks-MNAR			iks-MNAR		
	μ_q	$\mu_{\bar{q}}$		μ_q	$\mu_{\bar{q}}$		μ_q	$\mu_{\bar{q}}$
mc ₁₀	0.1	0.1	ks ₁₀	0	0.2	iks _{C1}	[.4, .5]	(0, .1]
mc ₂₀	0.2	0.2	ks ₂₀	0	0.4	iks _{C2}	[.3, .4]	[.1, .2]
mc ₃₀	0.3	0.3	ks ₃₀	0	0.6	iks _{C3}	[.2, .3]	[.2, .3]
mc ₄₀	0.4	0.4	ks ₄₀	0	0.8	iks _{C4}	[.1, .2]	[.3, .4]
						iks _{C5}	(0, .1]	[.4, .5]

Anmerkung. Die tiefgestellten Zahlen bei den MCAR und ks-MNAR Kombinationen geben den zu erwartenden Anteil fehlender Antworten an.

Für jede dieser 13 Simulationsarten³ wurden 200 Datensätze simuliert, die Stichprobengröße betrug dabei jeweils $N = 100000$. Diese großen simulierten Stichproben ermöglichen asymptotische Aussagen über die Modelle. Für eine Einschätzung bei kleineren Stichproben wurde N darüber hinaus auch variiert. Es wurden Datensätze für die folgenden Werte $N = \{100, 200, 400, 800, 1600, 3200, 6400, 12800, 25600\}$ simuliert.

5.2 Empirische Wissensstruktur

Außer der künstlich simulierten Wissensstruktur wurden auch drei empirische Wissensstrukturen auf Basis der Normierungstichprobe des Freiburger Persönlichkeitssinventars (Fahrenberg et al., 2010) hergeleitet. Das FPI-R ist ein deutschsprachiger Persönlichkeitstest zur Messung individueller Persönlichkeitsmerkmale. Es handelt sich dabei um ein faktorenanalytisch und item-metrisch begründetes Verfahren (Fahrenberg et al., 2010), welches im deutschsprachigen Raum weit verbreitet ist. Einsetzbar ab 16 Jahren soll die Persönlichkeit auf $10 + 2$ Skalen⁴ gemessen werden. Die insgesamt 138 Items werden dabei auf einer dichotomen Skala („Stimmt“ vs. „Stimmt nicht“) beantwortet. Es existiert eine geschlechts- und altersspezifische Normierungstichprobe aus dem Jahr 1999 (Fahrenberg et al., 2010) bestehend aus $N = 3740$ Probanden. Deren mittleres Alter betrug 45.84 Jahre ($SD = 17.66$) und 47% gaben

³Vier verschiedene MCAR, vier verschiedene ks-MNAR und fünf verschiedene iks-MNAR Bedingungen.

⁴Die Skalen des FPI-R lauten: Lebenszufriedenheit, Soziale Orientierung, Leistungsorientierung, Gehemmtheit, Erregbarkeit, Aggressivität, Beanspruchung, Körperliche Beschwerden, Gesundheitssorgen, Offenheit sowie die zwei Sekundärskalen Extraversion und Emotionalität im Sinne der Persönlichkeitstheorie Hans Jürgen Eysencks.

an weiblich zu sein. Für die Erstellung der Wissensstrukturen in R (R Core Team, 2020) wurden ähnlich zu dem Vorgehen in der Bachelorarbeit von Maurer (2019) die Skalen Erregbarkeit, Gehemmtheit und Gesundheitssorgen betrachtet und angelehnt an die Item Tree Analyse (Schrepp, 1999; Van Leeuwe, 1974) die Wissensstrukturen erstellt. Dazu wurden die R-Pakete „relations“ (Meyer & Hornik, 2019), „Rgraphviz“ (Hansen et al., 2020) und „pks“ (Heller & Wickelmaier, 2013) verwendet. Die einzelnen Items der drei Skalen können Anhang B entnommen werden.

Für die je 12 Items der Skalen Erregbarkeit, Gehemmtheit und Gesundheitssorgen⁵ wurden jeweils nur die kompletten Antwortmuster betrachtet. Da in der Normierungsstichprobe fehlende Antworten enthalten waren, reduzierte sich die Stichprobengröße je nach Skala etwas. In Tabelle 2 sind die absolute Häufigkeit fehlender Antworten sowie die Anzahl einzigartiger Antwortmuster für die drei Skalen zusammengefasst. Die Stichprobenkennwerte für die Stichproben der einzelnen Skalen veränderten sich aufgrund der fehlenden Daten nur marginal.

Tabelle 2

Absoluter Anteil fehlender Antworten (0, 1 oder 2), Anzahl verschiedener Antwortmuster sowie der gewählte Schwellenwert L für die drei Skalen des FPI-R.

Skala	Fehlende Antworten			Antwortmuster	Schwelle L
	0	1	2		
Erregbarkeit	3691	48	1	1240	194
Gehemmtheit	3679	57	4	1316	270
Gesundheitssorgen	3706	34	0	1350	277

Anmerkung. Aufgrund einiger Antwortmuster mit fehlenden Antworten reduziert sich die zugrunde liegende Stichprobe für jede Skala entsprechend.

Zur Untersuchung der Abhängigkeitsbeziehungen wurden alle paarweisen Kombinationen der Items $p, q \in Q$ für alle Skalen getrennt betrachtet. Die absolute Häufigkeit b_{pq} mit der Item p mit „Stimmt nicht“ und gleichzeitig Item q mit „Stimmt“ beantwortet wurde, stellt dann ein Maß dafür dar, wie stark für dieses Item die Precedence-Relation $p \preceq q$ (vgl. Einleitung) verletzt ist. Für die Erstellung einer Wissensstruktur muss nun eine Schwelle L definiert werden. Solange $b_{pq} \leq L$ gilt, wird eine Abhängigkeitsbeziehung zwischen den Items p und q , wie in der Precedence-Relation beschrieben, konstatiert und in die Struktur aufgenommen. Die

⁵Diese Skalen wurden aus den 10 + 2 Skalen des FPI-R ausgewählt, da sie zu moderat großen Wissensstrukturen bei „vergleichsweise guter Anpassung“ führen (J. Heller, Persönliche Kommunikation, 07.09.2020). Eine moderate Größe (500 - 1500) der Wissensstruktur war unumgänglich, da sonst der Rechenaufwand zu groß geworden wäre.

entsprechenden Schwellenwerte, die eine gute Anpassung bei moderater Größe der Wissensstruktur erzielten (J. Heller, Persönliche Kommunikation, 07.09.2020) sind Tabelle 2 zu entnehmen. Mit diesen wurden die Precedence-Relationen \preceq_{err} , \preceq_{geh} und \preceq_{ges} sowie darauf aufbauend die Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} bestimmt. Mit den absoluten Häufigkeiten für die verschiedenen Antwortmuster wurde nun das BLIM mit dem EM-Algorithmus an jede der drei Wissensstrukturen angepasst. Auf Basis der erhaltenen Parameter β , η und π wurden dann äquivalent zu dem Vorgehen bei der simulierten Wissensstruktur \mathcal{K}_s neue Datensätze mit fehlenden Daten simuliert. Auch hierbei wurden die drei verschiedenen Dezimierungsmechanismen MCAR, ks-MNAR und iks-MNAR angewendet. Aufgrund der Befunde in der Normierungsstichprobe wurde der Anteil fehlender Daten jedoch auf maximal 15% verringert, da bei den je 12 Items der Skalen maximal zwei fehlende ($\sim 17\%$) auftraten. Die entsprechenden Parameter für die verschiedenen Dezimierungsmechanismen sind Tabelle 3 zu entnehmen. Es wurden für die drei Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} in jeder Dezimierungsbedingung 200 Datensätze simuliert mit jeweils 3700 simulierten Antwortmustern⁶. Außerdem wurden die gleichen Simulationen mit 10000 simulierten Antwortmustern durchgeführt, um das Verhalten bei einer größeren Stichprobe zu betrachten. Die Ergebnisse für die größere simulierte Stichprobe mit $N = 10000$ sind Anhang F zu entnehmen.

Tabelle 3

Parameterwerte μ_q und $\mu_{\bar{q}}$ der verschiedenen Dezimierungsmechanismen für die simulierten Datensätze basierend auf den Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} des FPI-R.

	MCAR		ks-MNAR			iks-MNAR	
	μ_q	$\mu_{\bar{q}}$		μ_q	$\mu_{\bar{q}}$	μ_q	$\mu_{\bar{q}}$
mc ₁	0.01	0.01	ks ₁	0	0.02	iks _{c1}	[.10, .15] (0, .01]
mc ₅	0.05	0.05	ks ₅	0	0.10	iks _{c2}	[.05, .10] [.01, .05]
mc ₁₀	0.10	0.10	ks ₁₀	0	0.20	iks _{c3}	[.01, .05] [.05, .10]
mc ₁₅	0.15	0.15	ks ₁₅	0	0.30	iks _{c4}	(0, .01] [.10, .15]

Anmerkung. Die tiefgestellten Zahlen bei den MCAR und ks-MNAR Kombinationen geben den zu erwartenden Anteil fehlender Antworten an.

⁶Diese Zahl war angelegt an die Stichprobengröße der Normierungsstichprobe des FPI-R.

5.3 Implementierung der Modelle in R

Um die Modelle BLIM (BLIM_{MAW} bzw. BLIM_{COMP}), IMBLIM und MissBLIM an die simulierten Datensätze anzupassen, wurden diese in R (R Core Team, 2020) bzw. C++ (Stroustrup, 1998) mit Hilfe der Pakete „Rcpp“ (Eddelbuettel & François, 2011) und „RcppEigen“ (Bates & Eddelbuettel, 2013) implementiert. Die Zuhilfenahme einer kompilierten Programmiersprache wie C++ und der Vorlagenbibliothek „Eigen“ (Guennebaud & Jacob, 2010) für die lineare Algebra war nötig um die Rechenzeiten zu verkürzen und die CPU- und Arbeitsspeicherressourcen effizienter zu nutzen.

5.4 Modellvergleich

Um die verschiedenen Modelle zu vergleichen wurden äquivalent zu dem Vorgehen von De Chiusole et al. (2015) sowohl die Verzerrungen der Parameterschätzungen $\hat{\beta}_q - \beta_q^{true}$ und $\hat{\eta}_q - \eta_q^{true}$ berechnet als auch die Wiederherstellung des unterliegenden Wissenszustands K betrachtet. Die Verzerrungen wurden jeweils in Abhängigkeit des Modells (BLIM_{MAW}, BLIM_{COMP}, IMBLIM und MissBLIM) sowie des fehlende Daten generierenden Prozesses (MCAR, ks-MNAR und iks-MNAR) gemittelt und für einen Überblick wurde die mittlere Verzerrung über alle Aufgaben q berechnet. Für die Wiederherstellung des unterliegenden Wissenszustands K einer Person wurde der Modalwert der Wahrscheinlichkeiten $P(K | R, M, \hat{\theta})$ (vgl. Gleichung (12)) für alle $K \in \mathcal{K}$ berechnet und als der durch das Modell vorhergesagte unterliegende Wissenszustand \hat{K} angenommen. Um die Genauigkeit von \hat{K} zu bestimmen wurde die symmetrische Mengendistanz $d(\hat{K}, K^{true}) = |(\hat{K} \setminus K^{true}) \cup (K^{true} \setminus \hat{K})|$ für jedes an die Daten angepasste Modell berechnet. K^{true} ist dabei der wahre Wissenszustand der Person. Auch die symmetrische Mengendistanz wurde in Abhängigkeit des Modells und des fehlende Daten generierenden Prozesses (MCAR, ks-MNAR und iks-MNAR) gemittelt, um die Daten besser veranschaulichen zu können.

Zusätzlich zu den Auswertungen von De Chiusole et al. (2015) wurde auch die Streuung der Verzerrung für $\hat{\beta}_q$ und $\hat{\eta}_q$ für die verschiedenen Modelle betrachtet. Diese kann bei gleicher Verzerrung ein weiteres Kriterium für die Auswahl eines Modells darstellen.

6 Ergebnisse

Nachfolgend werden die Ergebnisse zuerst basierend auf der simulierten Wissensstruktur \mathcal{K}_s und anschließend basierend auf den empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} vorgestellt. Für die Abbildungen wurde das R-Paket „ggplot2“ (Wickham, 2016) verwendet.

6.1 Simulierte Wissensstruktur

Die simulierte Wissensstruktur \mathcal{K}_s umfasste 500 zufällige Wissenszustände $K \in \mathcal{K}_s$ und basierte auf einer Aufgabenmenge Q mit 25 Elementen. Die β - und η -Parameter wurden zufällig aus der Gleichverteilung $\text{unif}(0, 0.1)$ gezogen. Es gab keine im Vorhinein festgelegten Abhängigkeiten und es wurden auch keinerlei Veränderungen (z.B. Abgeschlossenheit bezüglich Mengenvereinigung) an der simulierten Struktur vorgenommen. Das führt dazu, dass die enthaltenen Abhängigkeiten implizit und womöglich sehr zahlreich bleiben.

6.1.1 Parameterwiederherstellung

In Tabelle 4 sind die Verzerrungen der Parameterschätzungen in Abhängigkeit des Modells und Dezimierungsmechanismus über die 25 Aufgaben gemittelt dargestellt. Abbildungen dieser Verzerrungen in Abhängigkeit der wahren Parameterwerte können Anhang C entnommen werden⁷. Es zeigt sich für den MCAR-Fall, dass das $\text{BLIM}_{\text{COMP}}$ bei hinreichend großer Stichprobe keine systematischen Verzerrungen aufweist. Dies ist dann der Fall, wenn aufgrund nicht zu hohem Anteil fehlender Antworten die Stichprobe mit kompletten Daten noch hinreichend groß ist. Tabelle 5 ist zu entnehmen, dass mit steigendem Anteil fehlender Daten, die Anzahl vollständiger Antwortmuster rapide zurückgeht. Das BLIM_{MAW} dagegen zeigt deutliche Verzerrungen, die den in Gleichung (7) und (8) bestimmten theoretischen entsprechen. Für IMBLIM und MissBLIM gilt, dass diese bei MCAR keine systematischen Verzerrungen aufweisen.

Für den ks-MNAR-Fall zeigen sich beim $\text{BLIM}_{\text{COMP}}$ ab 30% fehlender Daten systematische Verzerrungen. Das BLIM_{MAW} schätzt wie zu erwarten die β -Parameter genau, aber bei den η -Parametern ergeben sich die theoretisch zu erwarteten Verzerrungen. Das IMBLIM schneidet etwas schlechter als das $\text{BLIM}_{\text{COMP}}$ ab und das

⁷Die Ergebnisse in den Abbildungen gehen wenig über die in Tabelle 4 dargestellten Ergebnisse hinaus. Für eine Vergleichbarkeit mit den Ergebnissen von De Chiusole et al. (2015) wurden sie jedoch in Anhang C beigelegt.

Tabelle 4

Verzerrungen der Parameterschätzungen in Abhängigkeit des Modells und Dezimierungsmechanismus über alle Aufgaben q gemittelt.

	BLIM _{COMP}		BLIM _{MAW}		IMBLIM		MissBLIM	
	Bias β	Bias η	Bias β	Bias η	Bias β	Bias η	Bias β	Bias η
mc ₁₀	-0.02	0.00	94.64	-5.02	0.00	0.01	0.00	0.01
mc ₂₀	-0.02	-0.24	189.26	-10.07	-0.04	0.00	-0.04	0.00
mc ₃₀	0.55	1.52	283.88	-15.11	-0.02	0.00	-0.02	0.00
mc ₄₀	415.20	418.12	378.48	-20.15	-0.02	0.00	-0.02	0.00
ks ₁₀	0.00	0.10	0.00	-10.08	-0.46	0.55	0.00	-0.02
ks ₂₀	0.10	-0.34	0.00	-20.15	-2.07	3.55	0.00	-0.01
ks ₃₀	-1.11	63.88	-0.01	-30.21	-6.02	20.60	-0.01	0.00
ks ₄₀	-13.69	565.39	-0.01	-40.27	-44.45	463.16	0.00	0.01
iks _{C1}	0.50	-0.04	425.93	-2.53	6.00	-3.02	0.01	0.01
iks _{C2}	-1.03	-0.23	331.15	-7.62	2.42	-1.69	0.01	-0.01
iks _{C3}	-0.54	0.08	236.39	-12.59	0.03	0.00	0.02	-0.01
iks _{C4}	0.07	0.43	141.92	-17.63	-1.76	2.34	0.00	-0.01
iks _{C5}	-0.55	1.45	47.06	-22.58	-3.15	5.81	-0.01	0.02

Anmerkung. Die Verzerrung wurde mit 1000 multipliziert, um die Darstellung zu verbessern. Die simulierte Stichprobengröße pro Datensatz war $N = 100000$ mit je 200 Datensätzen pro Simulationsart.

MissBLIM schätzt die Parameter wieder am genauesten.

Für den iks-MNAR-Fall zeigt sich, dass das $\text{BLIM}_{\text{COMP}}$ die Parameter gut wiederherstellen kann. Das BLIM_{MAW} weist gerade bei höherem Anteil fehlender Antworten große Verzerrungen auf, die auch in diesem Fall den theoretisch vorhergesagten entsprechen. Das IMBLIM zeigt in diesen extremen Bereichen ebenfalls leichte Verzerrungen. Das MissBLIM schneidet dagegen auch hier sehr gut ab, es sind praktisch keine systematischen Verzerrungen vorhanden.

Tabelle 5

Mittlere Anzahl \bar{N} und relative Häufigkeit h vollständiger Antwortmuster für die Anpassung des $\text{BLIM}_{\text{COMP}}$ in Abhängigkeit des Dezimierungsmechanismus.

	\bar{N}	h		\bar{N}	h		\bar{N}	h
mc ₁₀	7174.64	0.0717	ks ₁₀	7275.51	0.0728	iks _{C1}	126.08	0.0013
mc ₂₀	377.01	0.0038	ks ₂₀	533.80	0.0053	iks _{C2}	79.33	0.0008
mc ₃₀	13.56	0.0001	ks ₃₀	168.71	0.0017	iks _{C3}	76.32	0.0008
mc ₄₀	1.15	0.0000	ks ₄₀	153.95	0.0015	iks _{C4}	77.11	0.0008
						iks _{C5}	121.74	0.0012

Anmerkung. Die Relative Häufigkeit h entspricht $\bar{N} \div 100000$.

Für das MissBLIM wurde ebenfalls die Wiederherstellung der μ - und $\bar{\mu}$ -Parameter betrachtet. Wie Tabelle 6 zu entnehmen ist, wurden diese Parameter ebenfalls sehr genau wiederhergestellt. Sie entsprechen praktisch den wahren Parametern, die der Datensimulation zugrunde lagen bzw. für iks-MNAR dem jeweiligen Intervallmittelpunkt der zugrundeliegenden Gleichverteilung.

Tabelle 6

Geschätzte μ - und $\bar{\mu}$ -Parameter für das MissBLIM in Abhängigkeit des Modells und Dezimierungsmechanismus über alle Aufgaben q gemittelt.

	MCAR		ks-MNAR		iks-MNAR			
	μ	$\bar{\mu}$	μ	$\bar{\mu}$	μ	$\bar{\mu}$		
mc ₁₀	0.100	0.100	ks ₁₀	0.000	0.200	iks _{C1}	0.450	0.050
mc ₂₀	0.200	0.200	ks ₂₀	0.000	0.400	iks _{C2}	0.350	0.151
mc ₃₀	0.300	0.300	ks ₃₀	0.000	0.600	iks _{C3}	0.250	0.250
mc ₄₀	0.400	0.400	ks ₄₀	0.000	0.800	iks _{C4}	0.150	0.350
						iks _{C5}	0.050	0.449

Neben der mittleren Verzerrung ist auch deren Streuung (vgl. Tabelle 7) von Interesse. Hierbei zeigte sich, dass für das $\text{BLIM}_{\text{COMP}}$ die Streuung mit steigendem Anteil fehlender Daten deutlich zunahm. Dies war bei allen drei Dezimierungsmecha-

nismen der Fall. Die Streuung der Verzerrung des BLIM_{MAW} war durchweg geringer, aber auch hierbei zeigten sich systematische Zusammenhänge: Mit steigendem Anteil fehlender Antworten, nahm die Streuung tendenziell zu. IMBLIM und MissBLIM zeigten Unterschiede bei hohem Anteil fehlender Daten. Das MissBLIM schnitt hier mit einer konstant geringen Streuung besser ab. Ansonsten waren diese beiden Modelle aber sehr ähnlich mit geringen Streuungen, die beim MissBLIM unabhängig vom Dezimierungsmechanismus waren.

Tabelle 7

Standardabweichung der Verzerrungen der Parameterschätzungen in Abhängigkeit des Modells und Dezimierungsmechanismus über alle Aufgaben q berechnet.

	$\text{BLIM}_{\text{COMP}}$		BLIM_{MAW}		IMBLIM		MissBLIM	
	SD β	SD η	SD β	SD η	SD β	SD η	SD β	SD η
mc_{10}	3.82	3.73	2.97	2.85	1.08	1.05	1.08	1.05
mc_{20}	16.70	16.25	5.31	5.46	1.21	1.15	1.21	1.15
mc_{30}	101.45	99.95	7.69	8.12	1.34	1.30	1.34	1.30
mc_{40}	499.24	500.09	10.14	10.82	1.61	1.57	1.62	1.58
ks_{10}	3.56	3.98	1.01	5.44	1.03	1.14	1.01	1.09
ks_{20}	11.49	18.95	1.03	10.74	1.27	1.97	1.03	1.26
ks_{30}	17.79	225.99	0.99	16.07	2.27	7.28	0.98	1.55
ks_{40}	29.22	439.91	1.01	21.42	15.28	33.76	1.01	2.21
iks_{C1}	43.75	22.46	29.33	3.38	2.76	1.62	1.39	1.02
iks_{C2}	40.49	33.00	28.26	4.68	1.74	1.36	1.30	1.12
iks_{C3}	37.87	36.55	27.79	7.00	1.34	1.32	1.25	1.23
iks_{C4}	34.37	39.96	27.55	9.55	1.33	1.75	1.13	1.31
iks_{C5}	23.70	44.35	27.20	12.14	1.54	2.85	1.05	1.34

Anmerkung. Die Standardabweichung der Verzerrung wurde mit 1000 multipliziert um die Darstellung zu verbessern. Die simulierte Stichprobengröße pro Datensatz war $N = 100000$ mit je 200 Datensätzen pro Simulationsart.

Parameterwiederherstellung und Stichprobengröße. Das Ergebnis für die großen Stichproben mit $N = 100000$ zeigt das asymptotische Verhalten der Modelle. Um diese Verzerrungen zusätzlich in Abhängigkeit kleinerer Stichprobengrößen vergleichen zu können, wurden die gleichen Simulationen basierend auf der artifiziellen Wissensstruktur \mathcal{K}_s mit entsprechend kleineren Stichprobengrößen ($N = \{100, 200, 400, 800, 1600, 3200, 6400, 12800, 25600\}$) durchgeführt. Es wurden hierbei für MCAR und ks-MNAR nur die Parameter betrachtet, die zu 10% bzw. 20% fehlenden Daten führen und für iks-MNAR nur die Kombinationen iks_{C2} und iks_{C4} . Die anderen

Werte waren hier aufgrund ihrer unrealistischen Höhe nicht von Interesse. Abbildung 1 zeigt die Verzerrung des β -Parameters der verschiedenen Modelle gemittelt über alle Aufgaben in Abhängigkeit der simulierten Stichprobengröße N und des Dezimierungsmechanismus. Zusätzlich sind die Standardabweichungen für die Verteilungen der Verzerrungen als Fehlerbalken eingezeichnet. Es ist offensichtlich, dass das MissBLIM und IMBLIM beide unabhängig von der Stichprobengröße und bei allen Bedingungen sehr gut abschneiden. Die Streuung der Verzerrung ist bei $N = 100$ bereits sehr klein und nimmt mit steigendem N weiter ab. Das BLIM_{COMP} schneidet in der mc₁₀ und ks₁₀ Bedingung ebenfalls gut ab. Hier zeigen sich für Stichprobengrößen unter 500 zwar recht große Streuungen, diese verkleinern sich jedoch mit zunehmender Stichprobengröße. Für die mc₂₀ und ks₂₀ Bedingung zeigt sich bei geringen Stichprobengrößen eine große Verzerrung und Streuung. Diese nehmen jedoch mit steigender Stichprobengröße beide ebenfalls ab. Bei iks_{C2} und iks_{C4} weist das BLIM_{COMP} eine starke Verzerrung und große Streuungen auf, die auch mit steigender Anzahl an Personen nicht schnell nachlässt. Am schlechtesten schneidet auch hier das BLIM_{MAW} ab. Es zeigen sich relativ unabhängig von der Stichprobengröße die theoretisch vorhergesagten Verzerrungen (vgl. Gleichung (7)) für den β -Parameter. Zwar ist die Streuung recht gering, doch aufgrund der systematischen Verzerrung, die nicht mit steigender Stichprobengröße schwindet, ist die Wiederherstellungsleistung der Parameter für das BLIM_{MAW} unzureichend. Die Verzerrungen für die η -Parameter in Abhängigkeit der Stichprobengröße N unterscheiden sich strukturell nicht groß von denen der β -Parameter. Eine entsprechende Abbildung ist daher Anhang D zu entnehmen.

6.1.2 Wiederherstellung des latenten Wissenszustandes

Um neben der Genauigkeit der Parameterschätzung auch abschätzen zu können, wie gut die verschiedenen Modelle weiterhin in der Lage sind trotz fehlender Daten die latenten Wissenszustände der Personen zu identifizieren, wurde die symmetrische Distanz $d(\hat{K}, K^{true}) = |(\hat{K} \setminus K^{true}) \cup (K^{true} \setminus \hat{K})|$ für jede Person in allen Datensätzen berechnet. Es gilt auch hier, dass die Vergleichbarkeit zwischen dem BLIM_{COMP} und den anderen Modellen eingeschränkt ist, da bei diesem aufgrund der ausgeschlossenen inkompletten Antwortmuster die Anzahl der betrachteten Fälle deutlich geringer ist (vgl. Tabelle 5). Daher werden diese Ergebnisse getrennt berichtet. Abbildung 2 zeigt die gemittelte symmetrische Distanz in Abhängigkeit der Modelle und Dezimierungsmechanismen für die auf Basis der artifiziellen Wissensstruktur \mathcal{K}_s simulierten

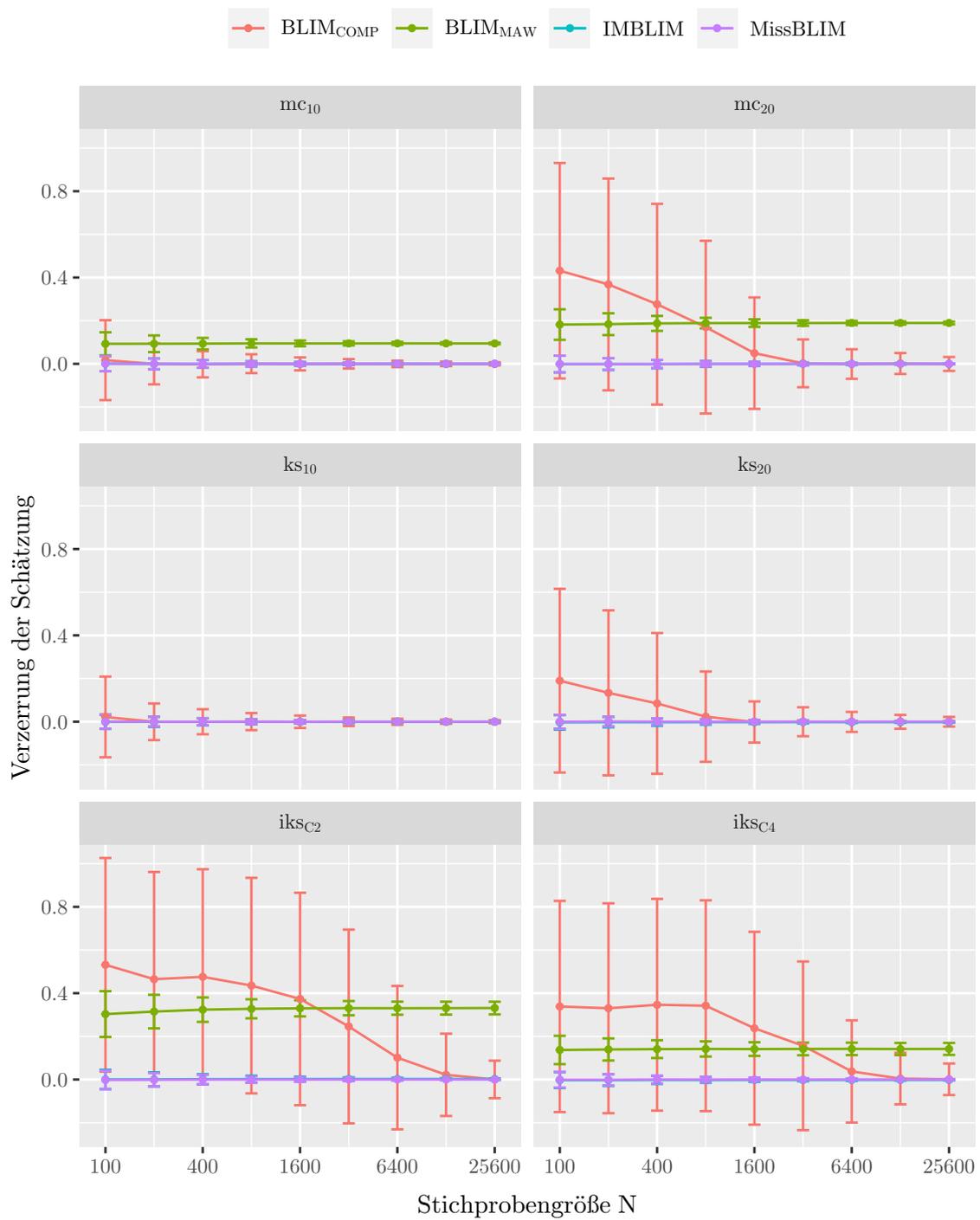


Abbildung 1. Verzerrung des β -Parameters der verschiedenen Modell gemittelt über alle Aufgaben in Abhängigkeit der simulierten Stichprobengröße N und des Dezimierungsmechanismus basierend auf der Wissensstruktur \mathcal{K}_s . Zusätzlich sind die Standardabweichungen als Fehlerbalken eingezeichnet. Die Abszisse ist zur Basis 2 logarithmiert.

Daten. Zusätzlich sind als Fehlerbalken die Standardabweichungen eingezeichnet. Es zeigt sich für die MCAR-Bedingung, dass das IMBLIM und MissBLIM fast identisch abschneiden. Die symmetrische Distanz vergrößert sich mit steigendem Anteil fehlender Daten, bleibt aber bis einschließlich 30% unterhalb 1. Das BLIM_{MAW} weist konstant deutlich höhere symmetrische Distanzen als IMBLIM und MissBLIM auf. Für den ks-MNAR Fall gilt, dass sowohl BLIM_{MAW} als auch MissBLIM sehr gut abschneiden. Die mittleren symmetrischen Distanzen liegen teils so nahe an 0, dass sie auf Abbildung 2 nicht zu erkennen sind. Gleiches gilt auch für deren Standardabweichungen. Das IMBLIM dagegen zeigt hier deutliche Defizite gegenüber den anderen Modellen: Mit über 6 wird bei 40% fehlenden Daten der höchste Wert erreicht. In der iks-MNAR Bedingung schneidet das MissBLIM am besten ab. Es zeigt für $C1$ und $C5$ sehr kleine Werte, die zwar zur Mitte ansteigen, aber unterhalb 0.5 bleiben. Die symmetrischen Distanzen beim IMBLIM sind sehr konstant auf dem höchsten Wert des MissBLIM. Das schlechteste Ergebnis liefert hier das BLIM_{MAW} mit sich verringernder symmetrischer Distanz mit absteigendem μ_q Parameter ($C1 - C5$).

Das $\text{BLIM}_{\text{COMP}}$ schnitt grundsätzlich sehr gut ab, wenn der Anteil fehlender Daten nicht zu groß war. In den Extrembereichen zeigten sich dann aber große symmetrische Distanzen und Verzerrungen, die vermutlich aufgrund der geringen Anzahl kompletter Antwortmuster zustande kamen (vgl. Tabelle 5).

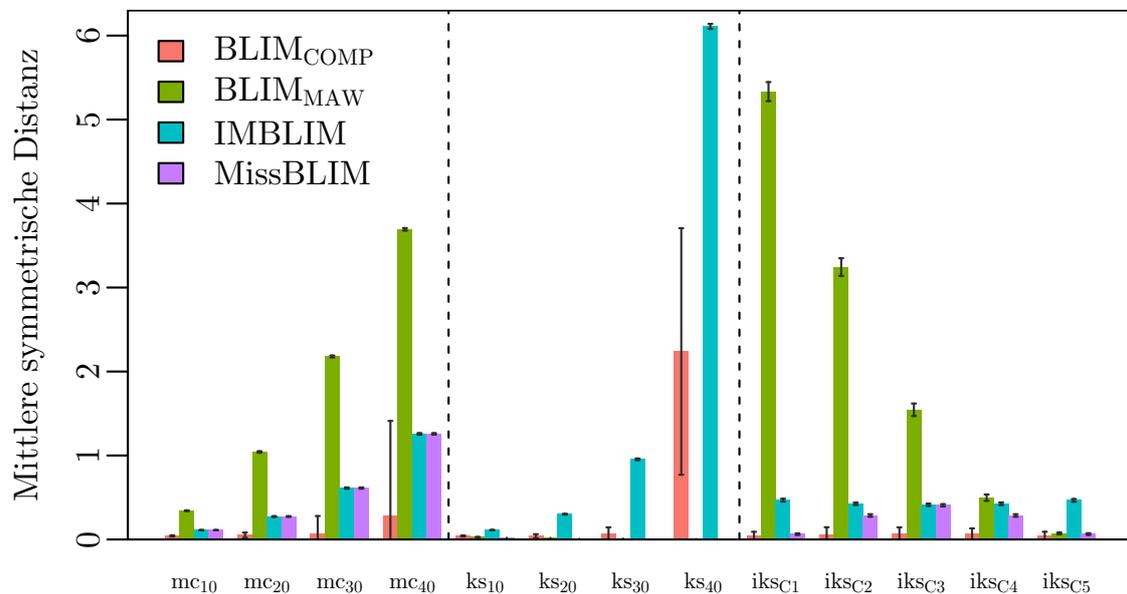


Abbildung 2. Mittlere symmetrische Distanz in Abhängigkeit des Modells und Dezimierungsmechanismus basierend auf der Wissensstruktur \mathcal{K}_s .

6.2 Empirische Wissensstrukturen

Die auf der Skala Erregbarkeit aus dem FPI-R (Fahrenberg et al., 2010) aufbauende Wissensstruktur \mathcal{K}_{err} umfasste 1120 Wissenszustände. Aufgrund der Methode der Erstellung war sie abgeschlossen bezüglich Mengenvereinigung und -durchschnitt und somit ein quasi-ordinaler Wissensraum. Die zugrunde liegende Precedence-Relation aus zwei Ebenen (vgl. Abbildung 3) zeigt, dass die Items 52, 93, 108 und 113 keinerlei Abhängigkeiten zu den anderen aufweisen und somit keinen Beitrag zu der Struktur leisten.

Für den Wissensraum auf der Skala Gehemtheit ergab sich eine Kardinalität von $|\mathcal{K}_{geh}| = 1024$, wobei die Items 4, 11, 73, 81, 120 und 124 keine Abhängigkeiten aufwiesen (vgl. Abbildung 4). Die zugehörige Precedence-Relation \preceq_{geh} bestand aus drei Ebenen.

Der letzte betrachtete Wissensraum auf der Skala Gesundheitssorgen umfasste 560 Wissenszustände und die Items 10 und 38 standen in keiner Abhängigkeit zu den anderen (vgl. Abbildung 5). Auch die Precedence-Relation \preceq_{ges} hatte drei Ebenen.

Alle drei Precedence-Relationen erfüllen die Eigenschaften Reflexivität ($\forall a \in Q : a \preceq a$), Transitivität ($\forall a, b, c \in Q : a \preceq b \wedge b \preceq c \implies a \preceq c$) und Antisymmetrie ($\forall a, b \in Q : a \preceq b \wedge b \preceq a \implies a = b$). Sie bilden daher je eine Halbordnung (I, \preceq) auf der zugehörigen Menge der Items I und man bezeichnet die korrespondierenden Wissensstrukturen dann auch als ordinale Wissensräume.

Identifizierbarkeit. Für die empirischen Wissensstrukturen wurde im Nachhinein aufgrund gewisser unerwarteter Effekte⁸ (vgl. Ergebnisse der β -Parameterschätzung bei Items 86 und 93 der Skala Erregbarkeit bei $N = 3700$) die Identifizierbarkeit der entsprechenden Parameter für das BLIM näher betrachtet, da vermutet wurde, dass diese Effekte eventuell dadurch bedingt sein könnten (J. Heller, Persönliche Kommunikation, 02.10.2020). Dabei zeigte sich, dass die Wissensstruktur \mathcal{K}_{err} auf der Skala Erregbarkeit in dem maximalen Item 60 backward-graded und in den minimalen Items 52, 86, 93, 108, 115 und 135 forward-graded war. Dies führt zu einer fehlenden Identifizierbarkeit der entsprechenden β bzw. η Parametern (Heller, 2017, Proposition 1 in Zusammenhang mit Corollary 1 i bzw. ii). Ebenso war die Wissensstruktur \mathcal{K}_{geh} im maximalen Item 97 backward-graded und in den minimalen Items 4, 11, 63, 73, 81 und 109 forward-graded. Zuletzt galt auch für \mathcal{K}_{ges} , dass diese im maximalen Item 127 backward-graded und in den minimalen Items 10, 38

⁸Diese relativierten sich jedoch bei der später durchgeführten Simulation mit größerer Stichprobe ($N = 10000$) wieder.

und 68 forward-graded war. Die möglichen Auswirkungen dieser fehlenden Identifizierbarkeit und die Betrachtung der Identifizierbarkeit im Zusammenhang mit dem IMBLIM und MissBLIM werden im folgenden nicht weiter betrachtet.

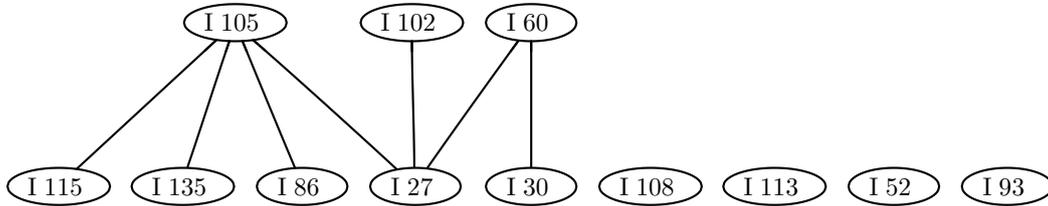


Abbildung 3. Hasse Diagramm der Precedence-Relation \preceq_{err} der Skala Erregbarkeit.

Nachfolgend sind für die drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} zuerst die Ergebnisse bei der Wiederherstellung der Parameter und anschließende der Wiederherstellung des latenten Wissenszustandes aufgeführt. Die Stichprobengröße der simulierten Daten war $N = 3700$. Die entsprechenden Abbildungen für eine Simulation mit einer größeren Stichprobe ($N = 10000$) können Anhang F entnommen werden. Diese zeigen geringere Schwankungen und für bestimmte Items (vgl. Items 86 und 93 der Skala Erregbarkeit) tendenziell eine bessere Schätzung, gehen ansonsten aber nicht über die berichteten Ergebnisse hinaus.

6.2.1 Parameterwiederherstellung

In den Abbildungen 6 bis 9 sind für das $\text{BLIM}_{\text{COMP}}$, BLIM_{MAW} , IMBLIM bzw. MissBLIM die β -Parameterschätzungen für die simulierten Daten aus den Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} in Abhängigkeit des wahren Parameterwertes für jedes Item und jeden Dezimierungsmechanismus dargestellt. Die entsprechenden Abbildungen für die Schätzungen der η -Parameter sind Anhang F zu entnehmen, zeigen aber für den Vergleich der Modelle sehr ähnliche Ergebnisse mit tendenziell geringeren Verzerrungen für die η -Parameter. Für die bessere Vergleichbarkeit mit den Ergebnissen der simulierten Wissensstruktur \mathcal{K}_s sind entsprechende Tabellen für die empirischen Wissensstrukturen Anhang E zu entnehmen.

Auffällig ist bei den Abbildungen 6 bis 9, dass viele der wahren Parameterwerte sehr nahe an 0 liegen. Außerdem scheinen bestimmte Items grundsätzlich bei allen Modellen eine hohe Verzerrung aufzuweisen, die sich paradoxerweise mit steigendem

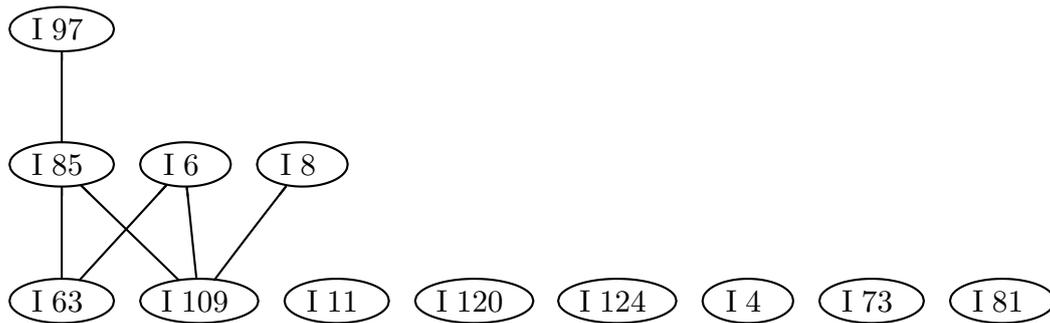


Abbildung 4. Hasse Diagramm der Precedence-Relation \preceq_{geh} der Skala Gehemtheit.

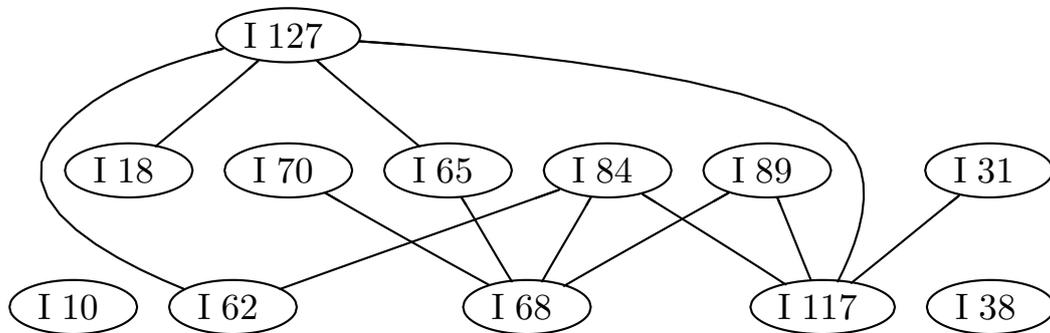


Abbildung 5. Hasse Diagramm der Precedence-Relation \preceq_{ges} der Skala Gesundheits-sorgen.

Anteil fehlender Daten teilweise verringert. Dies ist vor allem bei der Skala Erregbarkeit für die Items 86 und 93 mit wahren β -Parameterwerten von 0.067 bzw. 0.104 auffallend. Bei einer größeren simulierten Stichprobe von $N = 10000$ ist dieses Phänomen schwächer ausgeprägt, verschwindet jedoch nicht vollkommen (vgl. Anhang F).

Für das $\text{BLIM}_{\text{COMP}}$ zeigen sich bei kleinem Anteil fehlender Daten (1-5%) sehr kleine Verzerrungen und kleine Streuungen (vgl. Abbildung 6). Mit steigendem Anteil der fehlenden Daten nimmt diese Streuung und Verzerrung jedoch bei allen Dezimierungsmechanismen zu, was anhand der Abweichungen von der 1. Winkelhalbierenden und größeren Fehlerbalken zu erkennen ist. Diese Tendenz scheint bei MCAR etwas stärker ausgeprägt als bei ks-MNAR und iks-MNAR.

Das BLIM_{MAW} zeigt tendenziell die zu erwartenden theoretischen Verzerrungen (vgl. Abbildung 7). Dies gilt jedoch nicht, wenn der wahre Parameterwert 0 beträgt, da dann die Schätzung ebenfalls immer sehr nahe an 0 liegt und nicht die theoretisch zu erwartende Verzerrung auftritt. Wie schon beim $\text{BLIM}_{\text{COMP}}$ gilt, dass die Streuung mit steigendem Anteil fehlender Daten tendenziell ansteigt.

Beim IMBLIM zeigen sich generell geringere Verzerrungen, die Abweichungen in Abbildung 8 von der 1. Winkelhalbierenden sind geringer. Dieses Verhalten zeigt sich tendenziell über alle Dezimierungsmechanismen hinweg. Außerdem weist das IMBLIM hier eine geringere Verzerrungsanfälligkeit bei steigendem Anteil fehlender Daten auf und die Streuungen sind ebenfalls geringer.

Das MissBLIM schneidet noch etwas besser ab als das IMBLIM. Gerade bei höherem Anteil fehlender Daten sind die Streuungen geringer (vgl. Abbildung 9). Dies zeigt sich exemplarisch gut für den ks-MNAR Fall mit einem Anteil fehlender Daten von 15%: Hier ist Verzerrung und Streuung für das MissBLIM etwas geringer als für das IMBLIM.

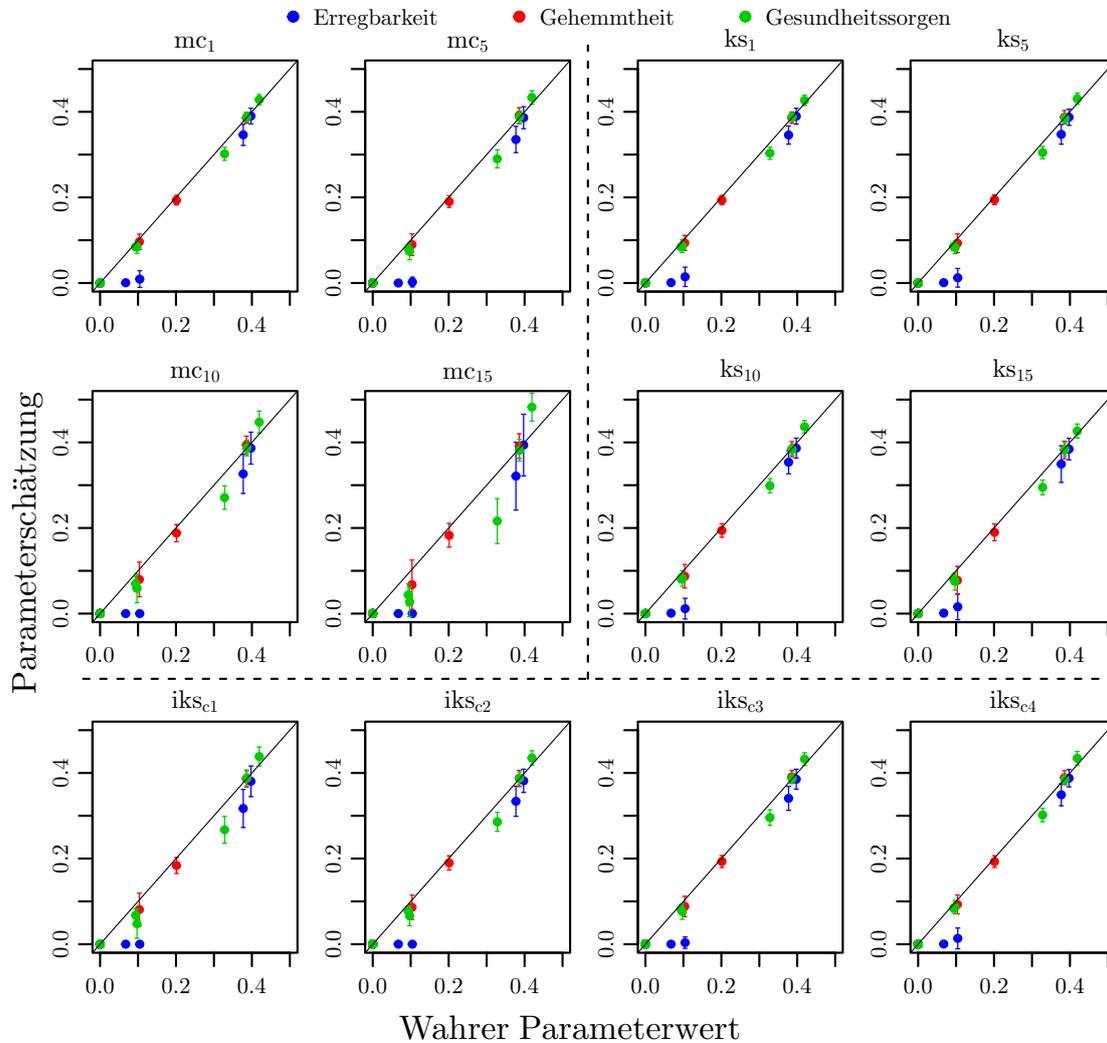


Abbildung 6. Mittlere β -Parameterschätzung des $BLIM_{COMP}$ in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} . Die durchgehende Linie zeigt die 1. Winkelhalbierende und die Standardabweichung ist als Fehlerbalken eingezeichnet.

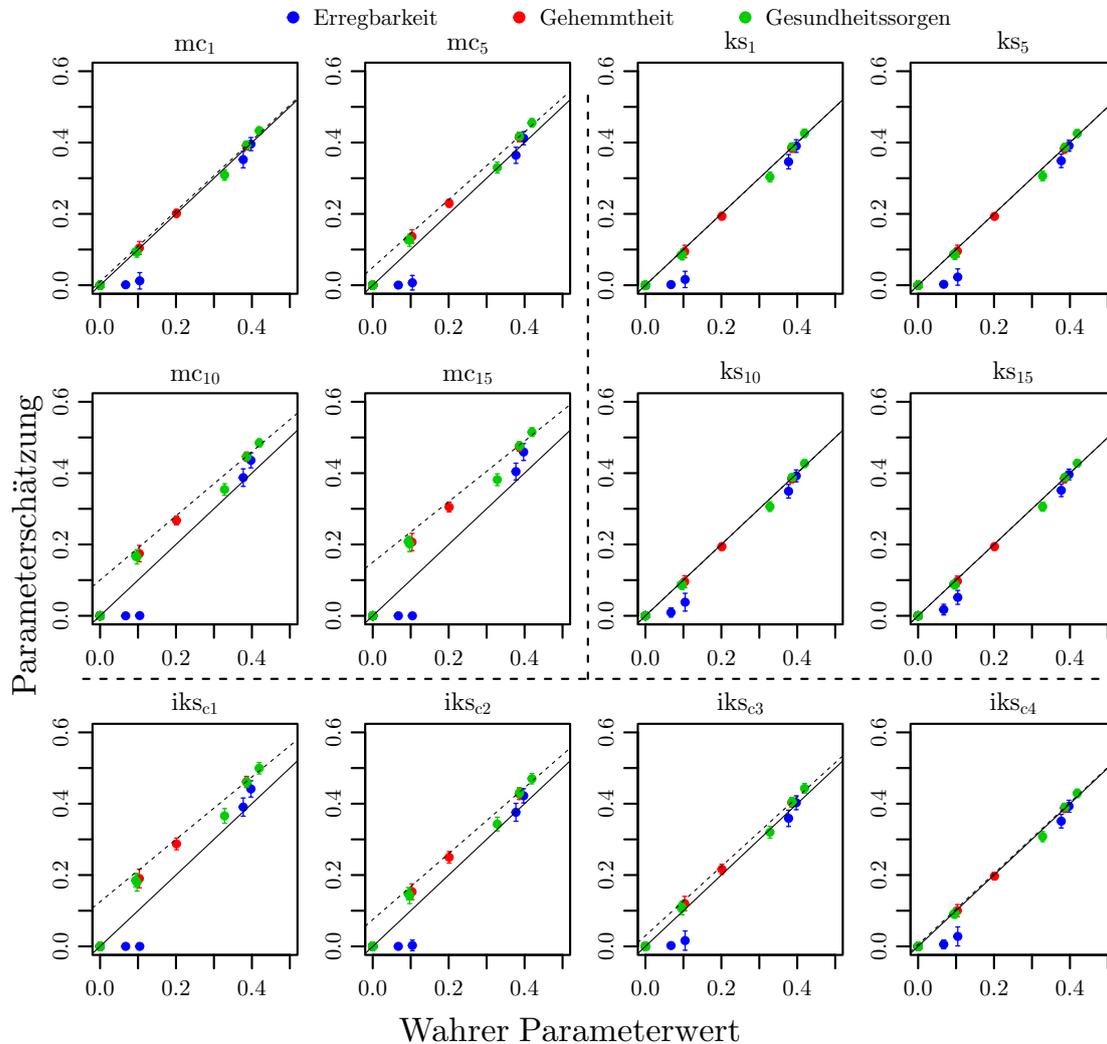


Abbildung 7. Mittlere β -Parameterschätzung des $BLIM_{MAW}$ in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} . Die durchgehende Linie zeigt die 1:1-Beziehung und die gestrichelte die theoretisch zu erwartende Verzerrung. Die Standardabweichung ist als Fehlerbalken eingezeichnet.

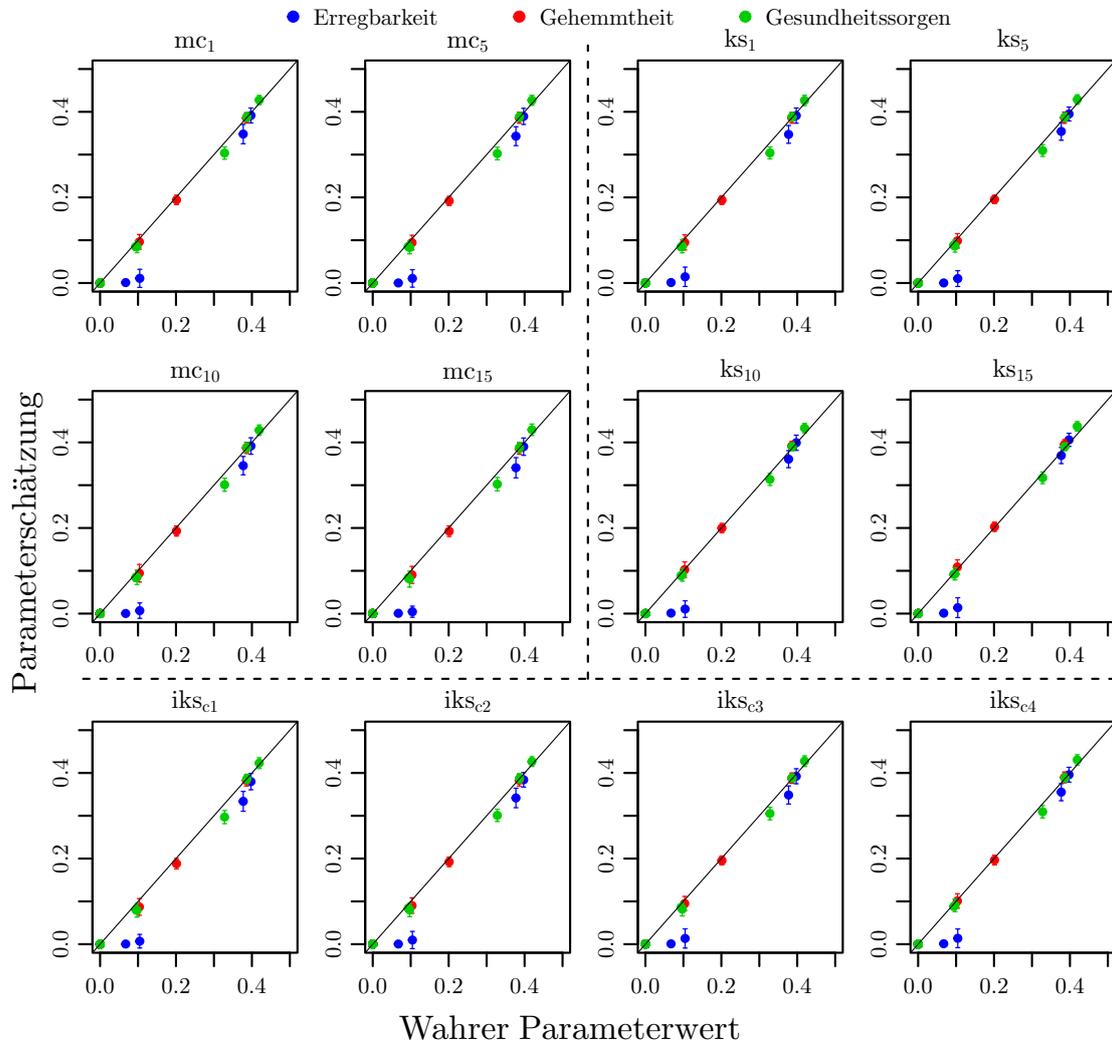


Abbildung 8. Mittlere β -Parameterschätzung des IMBLIM in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} . Die durchgehende Linie zeigt die 1. Winkelhalbierende und die Standardabweichung ist als Fehlerbalken eingezeichnet.

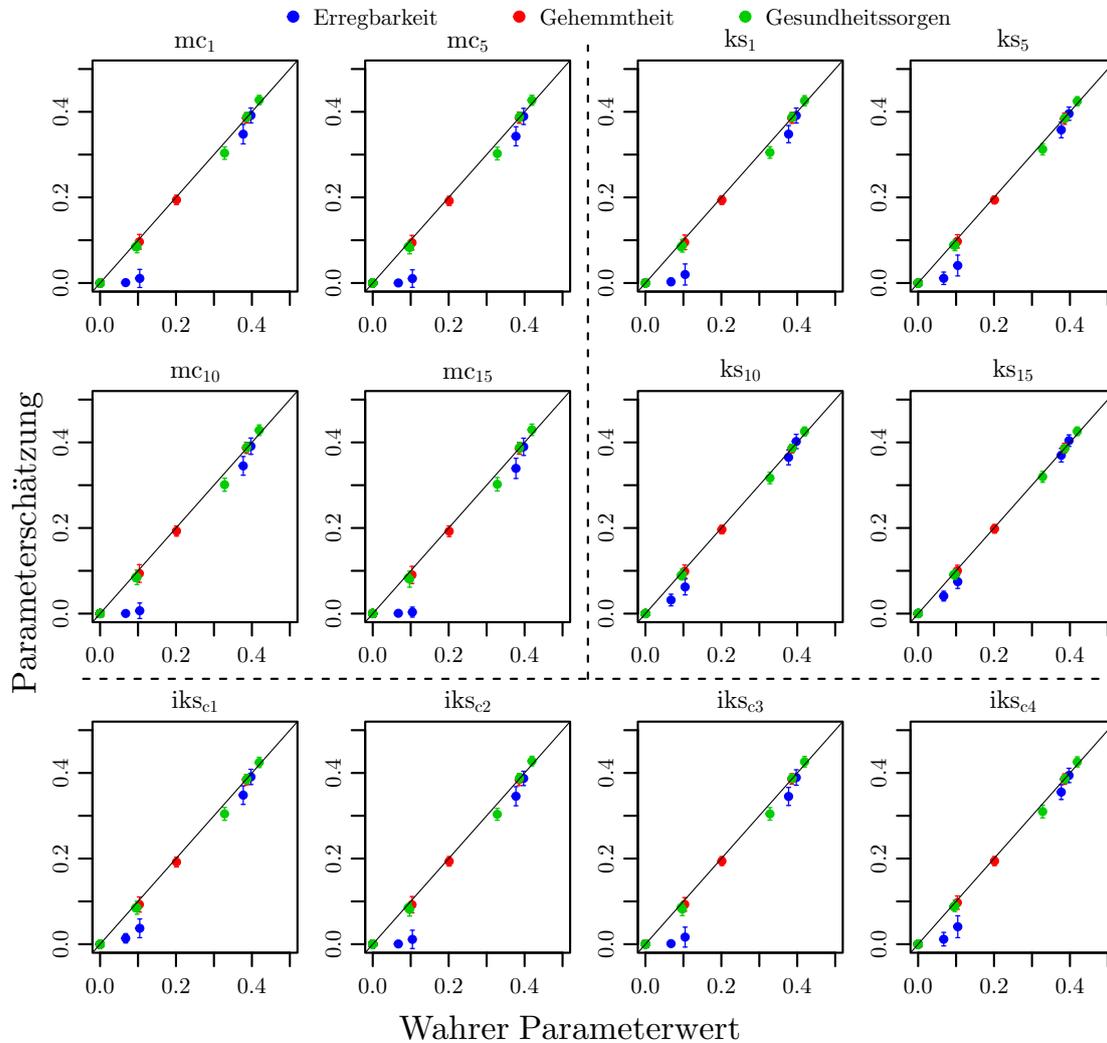


Abbildung 9. Mittlere β -Parameterschätzung des MissBLIM in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} . Die durchgehende Linie zeigt die 1:1-Beziehung und die Standardabweichung ist als Fehlerbalken eingezeichnet.

6.2.2 Wiederherstellung des latenten Wissenszustandes

Neben der Genauigkeit der Parameterschätzung spielt vor allem die Wiederherstellung des latenten Wissenszustandes einer Person bei der Diagnostik eine wichtige Rolle. Für diese zeigten sich bei den empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} ähnliche Ergebnisse wie schon bei der simulierten Wissensstruktur \mathcal{K}_s mit generell jedoch größeren Distanzen relativ zu dem kleineren Anteil fehlender Daten. Wie Abbildung 10 zu entnehmen ist, schnitt über alle drei empirischen Wissensstrukturen das $BLIM_{MAW}$ bei der MCAR und iks-MNAR Bedingung am schlechtesten ab. Es zeigte sich hier eine starke Abhängigkeit der symmetrischen Distanz von den entsprechenden μ - bzw. $\bar{\mu}$ -Parametern. Das IMBLIM schnitt bei diesen Bedingungen besser ab, zeigte aber in der ks-MNAR Bedingung größere Distanzen mit steigendem Anteil fehlender Daten. Das MissBLIM wies grundsätzlich mit die niedrigsten Distanzen auf, wurde jedoch in manchen Fällen gerade bei höherem Anteil fehlender Daten vom $BLIM_{COMP}$ übertroffen. Dies zeigte sich vor allem in der MCAR Bedingung. Auffällig war außerdem, dass in der ks-MNAR Bedingung die mittlere symmetrische Distanz mit steigenden $\bar{\mu}$ -Parametern für IMBLIM und MissBLIM sich unabhängig von der betrachteten Wissensstruktur zu verbessern schien. Beim $BLIM_{COMP}$ war das für die Skalen Gehemtheit und Gesundheitsorgen ebenso der Fall, zeigte sich bei der Skala Erregbarkeit jedoch eher umgekehrt.

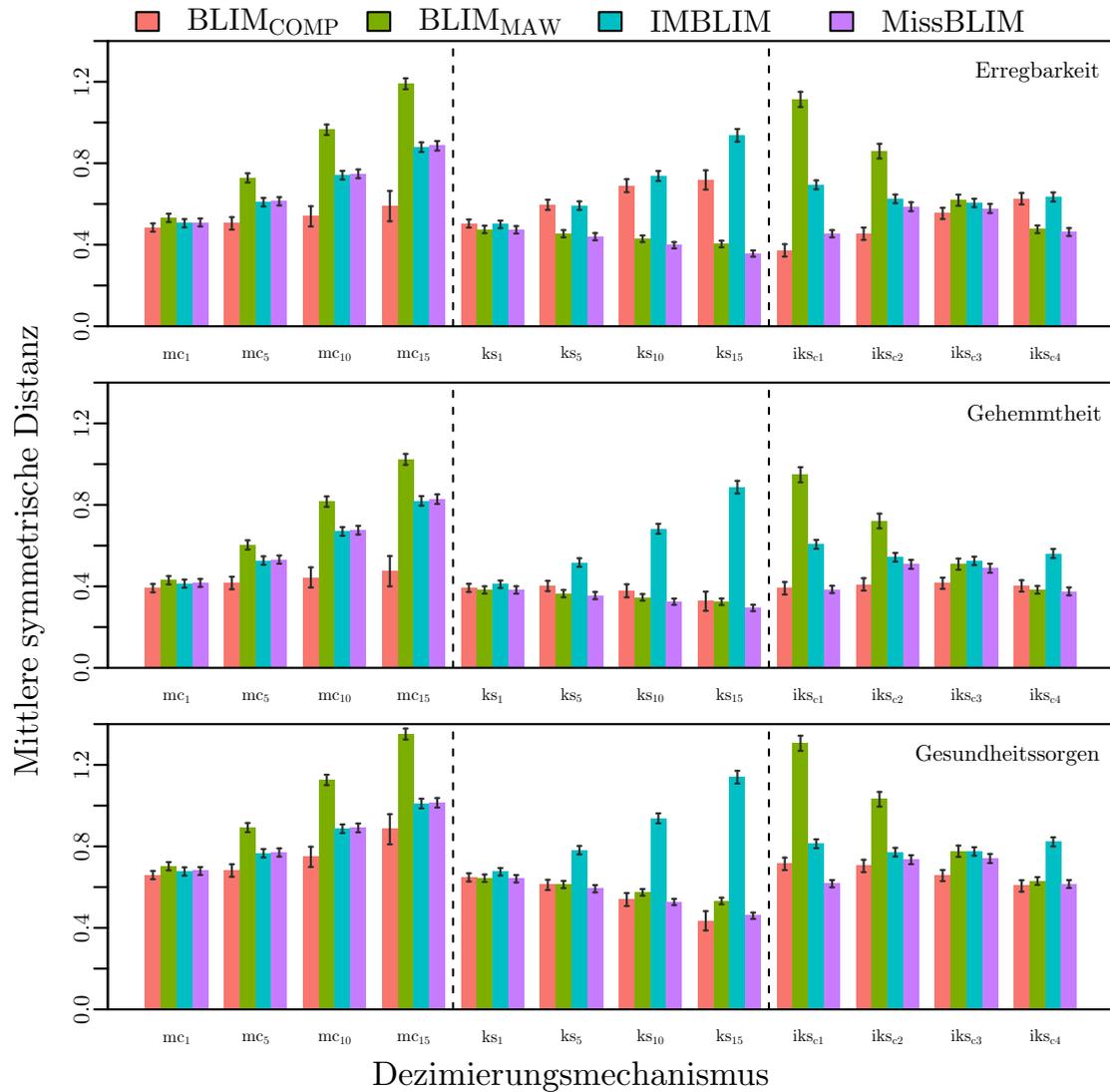


Abbildung 10. Mittlere symmetrische Distanz in Abhängigkeit des Modells und Dezimirungsmechanismus basierend auf den Skalen Erregbarkeit, Gehemmtheit und Gesundheitssorgen bzw. den unterliegenden Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine simulierte Stichprobengröße mit $N = 3700$.

7 Diskussion

Unterschiede zu De Chiusole et al. (2015). Für die Verzerrungen basierend auf der simulierten Wissensstruktur \mathcal{K}_s aus Tabelle 4 zeigten sich im Vergleich zu den Befunden von De Chiusole et al. (2015) Unterschiede. Diese scheinen vor allem bei hohen μ - und $\bar{\mu}$ -Parametern größer auszufallen. Leider waren der Studie keine Schätzgleichungen für die Anpassung der Modelle zu entnehmen. Jedoch haben die Autoren freundlicherweise ihren MATLAB-Code für die Schätzung des MissBLIM zur Verfügung gestellt. Aus diesem Code konnten die entsprechenden expliziten Schätzgleichungen $\hat{\beta}_q^{Padua}$ und $\hat{\eta}_q^{Padua}$, die im EM-Algorithmus Verwendung fanden, rekonstruiert werden (vgl. Gleichungen (18) und (19)).

$$\hat{\beta}_q^{Padua} = \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \setminus (R \cup M)}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K}} \quad (18)$$

$$\hat{\eta}_q^{Padua} = \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in R \setminus K}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in Q \setminus K}} \quad (19)$$

Verglichen mit den in Anhang A hergeleiteten ML-Schätzern für die β - und η -Parameter (Gleichungen (13) und (14)) zeigten sich hier Unterschiede im Nenner der Schätzgleichungen. In den ML-Schätzern der Originalstudie (De Chiusole et al., 2015) wurde vermutlich die Häufigkeit $F(R, M, K)$ auch dann addiert, wenn die Aufgabe q nicht beantwortet wurde, also $q \in M$ gilt. Dies scheint jedoch nicht korrekt, da die Häufigkeiten $F(R, M, K)$, für die gilt $q \in M$ nicht Teil der Summe im Nenner sein dürfen (vgl. Gleichungen (13) und (14)). Aufgrund dieser Unterschiede in den Schätzgleichungen ist anzunehmen, dass der Fehler für β_q bzw. η_q auch davon abhängt, wie oft gilt $i_{q \in K} \neq i_{q \in K \setminus M}$ bzw. $i_{q \in Q \setminus K} \neq i_{q \in Q \setminus (M \cup K)}$. Offensichtlich hängt das davon ab, wie wahrscheinlich eine Aufgabe $q \in K$ bzw. $q \notin K$ nicht beantwortet wird, also $q \in M$ gilt. Dies beschreiben gerade die Parameter μ_q bzw. $\mu_{\bar{q}}$, die im Dezimierungsmechanismus verändert werden. Somit ist auch nachvollziehbar, weshalb bei Veränderung dieser Parameter die Unterschiede zwischen den ML-Schätzern in der Originalstudie und den hergeleiteten ML-Schätzern variieren. Genau diese Unterschiede finden sich auch bei der Verzerrung der Schätzungen, wenn man Tabelle 4 dieser Arbeit mit Tabelle 1 von De Chiusole et al. (2015) vergleicht.

Modellvergleich. Der Vergleich der Modelle anhand der simulierten Wissensstruktur \mathcal{K}_s und der empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} basierend auf der Wiederherstellungsleistung der Parameter und des latenten Wissenszustandes zeigt,

dass je nach Prozess, der die fehlenden Daten generiert, die Modelle unterschiedlich gut abschneiden. In diesem Zusammenhang schnitt das BLIM_{MAW} am schlechtesten ab und zeigte die theoretisch zu erwartenden Verzerrungen in Abhängigkeit der μ - und $\bar{\mu}$ -Parameter (vgl. Gleichungen (7) und (8)). Es wies dadurch mit die höchsten Verzerrungen auf und auch deren Streuung war bei realistischem Anteil fehlender Daten unter 15% mit am größten. Dies war unabhängig von der Stichprobengröße der Fall. Bei der Wiederherstellungsleistung des latenten Wissenszustands zeigte es außer in der ks-MNAR Bedingung, hier ist die MAW-Annahme am wenigsten verletzt, da gilt $\mu_q = 0$ für alle $q \in Q$, die größten mittleren symmetrischen Distanzen, welche stark vom Anteil fehlender Daten abhingen.

Das $\text{BLIM}_{\text{COMP}}$ mit ausschließlich den vollständigen Antwortmustern schnitt besser ab als das BLIM_{MAW} . Dies liegt sehr wahrscheinlich daran, dass es keine Annahmen über den Prozess, der die fehlenden Daten generiert, macht, welche eventuell falsch sein könnten. Der Nachteil, der sich beim $\text{BLIM}_{\text{COMP}}$ jedoch gezeigt hat, ist die starke Abhängigkeit von der Stichprobengröße. Gerade bei hohem Anteil fehlender Daten nimmt die verbleibende Stichprobengröße rapide ab und es kommt dadurch zu teils großen Verzerrungen und Streuungen der Parameterschätzungen sowohl für β - als auch η -Parameter. Darüberhinaus ermöglicht es, wie schon das BLIM_{MAW} keinerlei Aussagen über einen möglichen Prozess, der den fehlenden Daten zugrunde liegen könnte. Es geht also mögliche Information, die auch in den fehlenden Daten enthalten sein kann, verloren.

Das IMBLIM wiederum ermöglicht es als erstes Modell die fehlenden Daten in die Schätzung der β - und η -Parameter mit einzubeziehen. Die Annahme, dass es eine über alle Aufgaben konstante Wahrscheinlichkeit gibt, mit der eine Antwort fehlt, egal ob die Aufgabe im Wissenszustand enthalten ist oder nicht, führt dazu, dass das IMBLIM die Parameter in der MCAR Bedingung sehr gut wiederherstellen kann. Auch ist hier die Streuung der Schätzung mit am geringsten, da in dieser Bedingung der Dezimierungsmechanismus genau der Annahme des IMBLIM entspricht. Für die ks-MNAR und iks-MNAR Bedingungen zeigt sich dann aber, dass eine Verzerrung auftritt, die mit steigendem Anteil fehlender Daten wächst. Ein ähnliches Muster zeigt sich dann auch bei der Wiederherstellungsleistung des latenten Wissenszustandes: Auch hier schneidet das IMBLIM im MCAR Fall sehr gut ab, weist aber gerade bei ks-MNAR und auch bei iks-MNAR größere mittlere symmetrische Distanzen, also eine ungenauere Herstellung des latenten Wissenszustandes auf.

Über alle Bedingungen hinweg zeigte das MissBLIM die besten Leistungen. Die Annahmen über den fehlende Daten generierenden Prozess sind flexibler als die des

IMBLIM und ermöglichen eine Abhängigkeit zwischen den fehlenden Antworten einer Person und deren unterliegendem Wissenszustand. Dies wird durch die zwei zusätzlichen Parameter μ_q und $\mu_{\bar{q}}$ für jede Aufgabe $q \in Q$ gewährleistet. μ_q ist dabei die Wahrscheinlichkeit, dass ein Item, welches im Wissenszustand K der Person enthalten ist ($q \in K$), nicht beantwortet wird und $\mu_{\bar{q}}$ entspricht der Wahrscheinlichkeit, dass die Antwort einer Person auf Aufgabe q fehlt, wenn diese nicht im Wissenszustand enthalten ist ($q \notin K$). Sowohl bei der simulierten Wissensstruktur \mathcal{K}_s ebenso wie bei den empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} hatte das MissBLIM tendenziell die geringsten Verzerrungen und kleinsten Streuungen, welche größtenteils unabhängig vom Dezimierungsmechanismus waren. Auch war es in der Lage die entsprechenden μ_q und $\mu_{\bar{q}}$ Parameter adäquat zu schätzen. Ein weiterer Vorteil gerade gegenüber dem $\text{BLIM}_{\text{COMP}}$ und BLIM_{MAW} war die recht genaue Schätzung auch bei kleineren Stichprobengrößen. Auch hier zeigten sich keine systematischen Verzerrungen, einzig die Streuung war etwas größer in diesen Bereichen, aber immer noch am kleinsten verglichen mit den anderen Modellen. Nicht zuletzt war auch die Wiederherstellungsleistung des latenten Wissenszustandes für das MissBLIM sehr zufriedenstellend. Es musste sich zwar teilweise gegenüber dem $\text{BLIM}_{\text{COMP}}$ geschlagen geben, dieses hat aber den bereits erwähnten Nachteil, dass bei kleineren Stichproben die Schätzgenauigkeit stark leidet und für Personen mit fehlenden Antworten kein Wissenszustand \hat{K} geschätzt werden kann. Einziger Nachteil des MissBLIM ist dagegen die größere Zahl der zu schätzenden Parameter und der damit einhergehende höhere Aufwand. Dieser Aufwand scheint aber gerechtfertigt, da durch ihn der Anteil verwertbarer Antwortmuster steigt. So müssen Datensätze mit fehlenden Antworten nicht ausgeschlossen werden, sondern deren Informationsgehalt kann mit in die Schätzung der Parameter einfließen und eine Schätzung des unterliegenden Wissenszustandes \hat{K} ist möglich. Dies ist ein wünschenswertes Vorgehen, welches auch gerade gegenüber den Klienten, Patienten und im wissenschaftlichen Kontext gegenüber den Versuchspersonen angebracht ist um deren Mühen und Bereitschaft Information zu liefern angemessen zu würdigen.

Diese Arbeit ermöglicht aufgrund der verwendeten Methodik eine eher theoretische Betrachtung der Modelle, welche zwar teils auf empirischen Wissensstrukturen aufbaut, die Anpassung der Modelle IMBLIM und MissBLIM selbst an empirische Datensätze jedoch nicht näher untersucht. In diesem Zusammenhang wäre daher eine Anpassung an einen empirischen Datensatz mit fehlenden Antworten, wie es De Chiusole et al. (2015) durchgeführt haben, zielführend. Gerade vor dem Hintergrund der

gefundenen Unterschiede in den ML-Schätzgleichungen wären diese Ergebnisse von Interesse. Hierbei wäre auch die Betrachtung verschiedener Ursachen fehlender Daten, z.B. Zeitbeschränkungen, bestimmte Instruktionen oder keinerlei Manipulation, und das entsprechende Abschneiden der Modelle in diesen Situationen spannend zu beobachten. Nicht zuletzt ist auch die Frage der Identifizierbarkeit für IMBLIM und MissBLIM nicht geklärt und es bleibt offen, inwieweit die zusätzlichen Parameter des MissBLIM hier ein Problem darstellen.

Literatur

- Bates, D. & Eddelbuettel, D. (2013). Fast and elegant numerical linear algebra using the RcppEigen package. *Journal of Statistical Software*, 52 (5), 1–24.
- De Chiusole, D., Stefanutti, L., Anselmi, P. & Robusto, E. (2015). Modeling missing data in knowledge space theory. *Psychological Methods*, 20 (4), 506–522. doi: 10.1037/met0000050
- Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39 (1), 1–22.
- Doignon, J. P. & Falmagne, J. C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, 23 (2), 175–196. doi: 10.1016/S0020-7373(85)80031-6
- Doignon, J.-P. & Falmagne, J.-C. (1999). *Knowledge spaces*. Berlin: Springer. doi: 10.1007/978-3-642-58625-5
- Eddelbuettel, D. & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40 (8), 1–18. doi: 10.18637/jss.v040.i08
- Fahrenberg, J., Hampel, R. & Selg, H. (2010). *Freiburger Persönlichkeitsinventar (FPI-R). 8., erweiterte Auflage. Manual*. Goettingen, Germany: Hogrefe.
- Guennebaud, G. & Jacob, B. (2010). *Eigen v3*. <http://eigen.tuxfamily.org>.
- Hansen, K. D., Gentry, J., Long, L., Gentleman, R., Falcon, S., Hahne, F. & Sarkar, D. (2020). Rgraphviz: Provides plotting capabilities for r graph objects [Software-Handbuch]. (R package version 2.32.0)
- Heller, J. (2017). Identifiability in probabilistic knowledge structures. *Journal of Mathematical Psychology*, 77, 46–57. doi: 10.1016/j.jmp.2016.07.008
- Heller, J. & Wickelmaier, F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics*, 42, 49–56. doi: 10.1016/j.endm.2013.05.145
- Little, R. J. A. & Rubin, D. B. (2002). *Statistical Analysis with Missing Data; 2nd ed.* Hoboken, NJ: Wiley.
- Maurer, A. (2019). *Careless responses in Fragebögen: Simulationen zur Detektion auf Grundlage probabilistischer Wissensstrukturen*. (Unveröffentlichte Bachelorarbeit). Universität Tübingen.
- Meyer, D. & Hornik, K. (2019). relations: Data structures and algorithms for relations [Software-Handbuch]. Zugriff auf <https://CRAN.R-project.org/package=relations> (R package version 0.6-9)

- R Core Team. (2020). R: A language and environment for statistical computing [Software-Handbuch]. Vienna, Austria. Zugriff auf <https://www.R-project.org/>
- Schrepp, M. (1999). On the empirical construction of implications between bi-valued test items. *Mathematical social sciences*, 38 (3), 361–375.
- Schrepp, M. (2005). About the connection between knowledge structures and latent class models. *Methodology*, 1 (3), 93–103.
- Stefanutti, L. & Robusto, E. (2009). Recovering a probabilistic knowledge structure by constraining its parameter space. *Psychometrika*, 74 (1), 83–96. doi: 10.1007/s11336-008-9095-7
- Stroustrup, B. (1998). An Overview of the C++ Programming Language. doi: 10.1201/9780849331350.sec3
- Ünlü, A. (2011). A note on the connection between knowledge structures and latent class models. *Methodology*.
- Van Leeuwe, J. F. (1974). Item tree analysis. *Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden*, 29 (6), 475–483.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Zugriff auf <https://ggplot2.tidyverse.org>

A Herleitung ML-Schätzer

Herleitung der ML-Schätzer für das BLIM, IMBLIM und MissBLIM Modell.

Für den M-Schritt des EM-Algorithmus ist eine ML-Schätzung der Parameter durchzuführen. Nachfolgend sollen die expliziten Schätzgleichungen für das IMBLIM und MissBLIM im Durchgang⁹ t hergeleitet werden. $F(R, M, K)$ bezieht sich dabei auf die im E-Schritt berechneten erwarteten Häufigkeiten $\mathcal{E}(F(R, M, K))$ (vgl. Gleichung (11)). Die Parameterschätzungen des BLIM entsprechen aufgrund der Beziehung zwischen BLIM und IMBLIM denen des IMBLIM mit $M = \emptyset$.

IMBLIM

Ausgehend von der Likelihood der vollständigen Daten (9) wird die log-Likelihood gebildet:

$$\begin{aligned}
 \log \mathcal{L}(\mathcal{Q}_K \mid \beta, \eta, \pi) &= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log (P(M) \cdot P(\langle R, M \rangle^* \mid K, \beta, \eta)) \\
 &+ \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log \pi_K \\
 &= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log P(M) \\
 &+ \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log P(\langle R, M \rangle^* \mid K, \beta, \eta) \\
 &+ \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log \pi_K
 \end{aligned}$$

\mathcal{Q}_K ist dabei die Menge aller theoretisch möglichen Tripel $\langle R, M, K \rangle$. \mathcal{Q} bezeichne weiter die Menge aller beobachtbaren Paare $\langle R, M \rangle$.

Für die ML-Schätzung muss diese log-Likelihood nun partiell nach den Parametern abgeleitet und 0 gesetzt werden. Hierzu ist hilfreich, dass

$$\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log P(\langle R, M \rangle^* \mid K)$$

nur von β und η abhängt und

$$\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log \pi_K$$

⁹Im Anhang wird auf eine Kennzeichnung des Durchgangs mit hochgestelltem t verzichtet, alle Schätzungen beziehen sich jedoch auf den Durchgang t .

nur von π_K . Somit fallen bei der partiellen Ableitung die jeweils anderen Summanden weg, da sie von dem abzuleitenden Parameter unabhängig sind.

π -Parameter

Hier ist der Teil $\log \mathcal{L}_\pi = \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log \pi_K$ unter der Nebenbedingung $\sum_{K \in \mathcal{K}} \pi_K \stackrel{!}{=} 1$ zu maximieren. Dies kann durch die Einführung eines Lagrange Multiplikators λ erreicht werden:

$$\begin{aligned} \log \mathcal{L}_\pi &= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log \pi_K \\ \log \mathcal{L}_\pi^{\text{Lagrange}} &= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log \pi_K + \lambda \cdot \left(1 - \sum_{K \in \mathcal{K}} \pi_K\right) \\ \frac{\partial \log \mathcal{L}_\pi^{\text{Lagrange}}}{\partial \pi_K} &= \sum_{\langle R, M \rangle \in \mathcal{Q}} \frac{F(R, M, K)}{\pi_K} - \lambda \end{aligned}$$

Nullsetzten und auflösen:

$$\hat{\pi}_K = \frac{1}{\lambda} \sum_{\langle R, M \rangle \in \mathcal{Q}} F(R, M, K)$$

Wegen

$$1 = \sum_{K \in \mathcal{K}} \hat{\pi}_K = \sum_{K \in \mathcal{K}} \frac{\sum_{\langle R, M \rangle \in \mathcal{Q}} F(R, M, K)}{\lambda} = \frac{N}{\lambda}$$

erhält man dann

$$\hat{\pi}_K = \frac{1}{N} \sum_{\langle R, M \rangle \in \mathcal{Q}} F(R, M, K)$$

wobei $N = \sum_{\langle R, M \rangle \in \mathcal{Q}} \sum_{K \in \mathcal{K}} F(R, M, K) = \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K)$ der Stichprobengröße entspricht.

β -Parameter

Hierzu ist der von β abhängige Teil der log-Likelihood $\log \mathcal{L}_\beta$ zu maximieren mit der bedingten Wahrscheinlichkeit $P(\langle R, M \rangle^* | K, \beta, \eta)$ wie in Gleichung (4) definiert.

$$\begin{aligned}
\log \mathcal{L}_\beta &= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log P(\langle R, M \rangle^* | K, \beta, \eta) \\
&= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in K \setminus (R \cup M)} \log \beta_q \\
&\quad + \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in (K \cap R) \setminus M} \log(1 - \beta_q) \\
&\quad + \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in R \setminus (K \cup M)} \log \eta_q \\
&\quad + \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in \mathcal{Q} \setminus (R \cup M \cup K)} \log(1 - \eta_q) \\
\frac{\partial \log \mathcal{L}_\beta}{\partial \beta_q} &= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in K \setminus (R \cup M)} \frac{1}{\beta_q} \\
&\quad + \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in (K \cap R) \setminus M} \frac{1}{\beta_q - 1}
\end{aligned}$$

Einführung von zwei Indikatorvariablen nach dem Schema wie in Gleichung (2):

$$\begin{aligned}
i_{q \in K \setminus (R \cup M)} &= \begin{cases} 1 & , \text{wenn } q \in K \setminus (R \cup M) \\ 0 & \text{sonst} \end{cases} \\
i_{q \in (K \cap R) \setminus M} &= \begin{cases} 1 & , \text{wenn } q \in (K \cap R) \setminus M \\ 0 & \text{sonst} \end{cases}
\end{aligned}$$

Daraus folgt:

$$\begin{aligned}
\frac{\partial \log \mathcal{L}_\beta}{\partial \beta_q} &= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \setminus (R \cup M)} \cdot \frac{1}{\beta_q} \\
&\quad + \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in (K \cap R) \setminus M} \cdot \frac{1}{\beta_q - 1}
\end{aligned}$$

Nullsetzen:

$$\begin{aligned}
0 &= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \setminus (R \cup M)} \cdot \frac{1}{\hat{\beta}_q} \\
&\quad + \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in (K \cap R) \setminus M} \cdot \frac{1}{\hat{\beta}_q - 1} \\
0 &= (\hat{\beta}_q - 1) \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \setminus (R \cup M)} \\
&\quad + \hat{\beta}_q \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in (K \cap R) \setminus M} \\
\hat{\beta}_q &= \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \setminus (R \cup M)}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \setminus (R \cup M)} + \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in (K \cap R) \setminus M}} \\
&= \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \setminus (R \cup M)}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot (i_{q \in K \setminus (R \cup M)} + i_{q \in (K \cap R) \setminus M})} \\
&= \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \setminus (R \cup M)}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \setminus M}}
\end{aligned}$$

η -Parameter

Das Vorgehen ist aufgrund der sehr ähnlichen Struktur der Gleichung (4) für β und η äquivalent und führt zu folgender Schätzgleichung mit angepassten Indikatorvariablen $i_{q \in R \setminus (K \cup M)}$ und $i_{q \in Q \setminus (R \cup M \cup K)}$, welche sich aufgrund $R \cap M = \emptyset$ vereinfachen lassen:

$$\begin{aligned}
\hat{\eta}_q &= \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in R \setminus (K \cup M)}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot (i_{q \in R \setminus (K \cup M)} + i_{q \in Q \setminus (R \cup M \cup K)})} \\
&= \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in R \setminus (K \cup M)}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in Q \setminus (M \cup K)}}
\end{aligned}$$

MissBLIM

Wieder ausgehend von der Likelihood der vollständigen Daten (10) diesmal des MissBLIM wird die log-Likelihood gebildet:

$$\begin{aligned}
\log \mathcal{L}(\mathcal{Q}_K | \beta, \eta, \pi, \mu, \bar{\mu}) &= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log (P(\langle R, M \rangle^* | K, \beta, \eta) \\
&\quad \cdot P(M | K, \mu, \bar{\mu})) \\
&\quad + \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log \pi_K \\
&= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log P(\langle R, M \rangle^* | K, \beta, \eta) \\
&\quad + \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log P(M | K, \mu, \bar{\mu}) \\
&\quad + \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log \pi_K
\end{aligned}$$

Die Schätzer für β , η und π sind dabei äquivalent zu denen des IMBLIM, da der von ihnen abhängige Teil der Likelihood gleich ist. Hinzu kommen hier jedoch noch die $\mu = (\mu_q)_{q \in Q}$ und $\bar{\mu} = (\mu_{\bar{q}})_{q \in Q}$ Parameter: μ_q als die Wahrscheinlichkeit, dass ein Item q , das im Wissenszustand K enthalten ist, nicht beantwortet wird (weder richtig, noch falsch) und $\mu_{\bar{q}}$, die Wahrscheinlichkeit, dass ein Item, das sich nicht im Wissenszustand der Person befindet, nicht beantwortet wird. Für die ML-Schätzung dieser Parameter muss der von diesen abhängige Teil der Log-Likelihood maximiert werden.

μ - und $\bar{\mu}$ -Parameter

Der von μ und $\bar{\mu}$ abhängige Teil der log-Likelihood lautet:

$$\log \mathcal{L}_\mu = \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \log P(M | K, \mu, \bar{\mu})$$

mit $P(M | K, \mu, \bar{\mu})$ wie in Gleichung (6) definiert. Für das einfachere Ableiten kann diese Gleichung umgeschrieben werden (vgl. De Chiusole et al., 2015):

$$P(M | K, \mu, \bar{\mu}) = \left[\prod_{q \in K \cap M} \mu_q \right] \left[\prod_{q \in K \setminus M} (1 - \mu_q) \right] \left[\prod_{q \in M \setminus K} \mu_{\bar{q}} \right] \left[\prod_{q \in Q \setminus (M \cup K)} (1 - \mu_{\bar{q}}) \right]$$

Damit ergibt sich dann für die log-Likelihood und die Ableitung

$$\begin{aligned}
\log \mathcal{L}_\mu &= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in K \cap M} \log \mu_q \\
&+ \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in K \setminus M} \log(1 - \mu_q) \\
&+ \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in M \setminus K} \log \mu_{\bar{q}} \\
&+ \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in Q \setminus (M \cup K)} \log(1 - \mu_{\bar{q}}) \\
\frac{\partial \log \mathcal{L}_\mu}{\partial \mu_q} &= \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in K \cap M} \frac{1}{\mu_q} \\
&+ \sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot \sum_{q \in K \setminus M} \frac{1}{\mu_q - 1}
\end{aligned}$$

Das weitere Vorgehen ist dann sehr ähnlich zu der β - bzw. η -Schätzung des IMBLIM, da die Gleichungen strukturell äquivalent sind. Es werden wieder zwei Indikatorvariablen zur Hilfe genommen:

$$\begin{aligned}
i_{q \in K \cap M} &= \begin{cases} 1 & , \text{wenn } q \in K \cap M \\ 0 & \text{sonst} \end{cases} \\
i_{q \in K \setminus M} &= \begin{cases} 1 & , \text{wenn } q \in K \setminus M \\ 0 & \text{sonst} \end{cases}
\end{aligned}$$

Damit ergibt sich dann die ML-Schätzung für μ_q :

$$\begin{aligned}
\hat{\mu}_q &= \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \cap M}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot (i_{q \in K \cap M} + i_{q \in K \setminus M})} \\
&= \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K \cap M}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in K}}
\end{aligned}$$

und für $\mu_{\bar{q}}$:

$$\begin{aligned}
\hat{\mu}_{\bar{q}} &= \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in M \setminus K}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot (i_{q \in M \setminus K} + i_{q \in Q \setminus (M \cup K)})} \\
&= \frac{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in M \setminus K}}{\sum_{\langle R, M, K \rangle \in \mathcal{Q}_K} F(R, M, K) \cdot i_{q \in Q \setminus K}}
\end{aligned}$$

B Items FPI-R

Die verwendeten Items des FPI-R für die Skalen lauten:

Tabelle B1

12 Items der Skala Gesundheitsorgen.

Skala	Item	Aussage
Gesundheitsorgen	10	Im Krankheitsfall möchte ich Befund und Behandlung eigentlich von einem zweiten Arzt überprüfen lassen.
	18	Ich achte aus gesundheitlichen Gründen auf regelmäßige Mahlzeiten und reichlichen Schlaf.
	31	Ich habe mich über die häufigsten Krankheiten und ihre ersten Anzeichen informiert.
	38	Um gesund zu bleiben, achte ich auf ein ruhiges Leben.
	62	Ich vermeide es, ungewaschenes Obst zu essen.
	65	Ich vermeide Zugluft, weil man sich zu leicht erkälten kann.
	68	Wenn jemand in meine Richtung hustet oder niest, versuche ich mich abzuwenden.
	70	Ich hole sicherheitshalber ärztlichen Rat ein, wenn ich länger als zwei Tage erhöhte Temperatur (leichtes Fieber) habe.
	84	Weil man sich so leicht anstecken kann, wasche ich mir zu Hause gleich die Hände.
	89	Ich passe auf, dass ich nicht zu viel Autoabgase und Staub einatme.
	117	Handtücher in viel benutzten Waschräumen sind mir wegen der Ansteckungsgefahr unangenehm.
	127	Auch ohne ernste Beschwerden gehe ich regelmäßig zum Arzt, nur zur Vorsicht.

Anmerkung. Negativ gepolte Items sind mit (n) gekennzeichnet.

Tabelle B2

Je 12 Items der Skalen Erregbarkeit und Gehemmtheit.

Skala	Item	Aussage
Erregbarkeit	27	Ich neige dazu, bei Auseinandersetzungen lauter zu sprechen als sonst.
	30	Wenn mir einmal etwas schiefgeht, regt mich das nicht weiter auf. (n)
	52	Auch wenn es eher viel zu tun gibt, lasse ich mich nicht hetzen. (n)
	60	Auch wenn mich etwas sehr aus der Fassung bringt, beruhige ich mich meistens wieder rasch. (n)
	86	Mein Blut kocht, wenn man mich zum Narren hält.
	93	Es gibt nur wenige Dinge, die mich leicht erregen oder ärgern. (n)
	102	Im allgemeinen bin ich ruhig und nicht leicht aufzuregen. (n)
	105	Ich kann oft meinen Ärger und meine Wut nicht beherrschen.
	108	Ich lasse mich durch eine Vielzahl von kleinen Störungen nicht aus der Ruhe bringen. (n)
	113	Ich neige oft zu Hast und Eile, auch wenn es überhaupt nicht notwendig ist.
	115	Oft rege ich mich zu rasch über jemanden auf.
	135	Ich bin leicht aus der Ruhe gebracht, wenn ich angegriffen werde.
Gehemmtheit	4	Ich habe fast immer eine schlagfertige Antwort bereit. (n)
	6	Ich scheue mich, allein in einen Raum zu gehen, in dem andere Leute bereits zusammensitzen und sich unterhalten.
	8	Ich würde mich beim Kellner oder Geschäftsführer eines Restaurants beschweren, wenn ein schlechtes Essen serviert wird. (n)
	11	Ich bin ungern mit Menschen zusammen, die ich noch nicht kenne.
	63	Es fällt mir schwer, vor einer großen Gruppe von Menschen zu sprechen oder vorzutragen.
	73	Ich bin im Grunde eher ein ängstlicher Mensch.
	81	Ich schließe nur langsam Freundschaften.
	85	Ich werde ziemlich leicht verlegen.
	97	Ich erröte leicht.
	109	Bei Geselligkeiten und öffentlichen Veranstaltungen bleibe ich lieber im Hintergrund.
	120	Beim Reisen schaue ich lieber auf die Landschaft als mich mit den Mitreisenden zu unterhalten.
	124	Es fällt mir schwer, den richtigen Gesprächsstoff zu finden, wenn ich jemanden kennenlernen will.

Anmerkung. Negativ gepolte Items sind mit (n) gekennzeichnet.

C Abbildungen simulierte Wissensstruktur

Nachfolgend sind für die simulierte Wissensstruktur \mathcal{K}_s die Parameterschätzungen in Abhängigkeit der wahren Parameterwerte abgebildet. Dies geschieht für jedes Modell und jeden Dezimierungsmechanismus. Bei keiner Verzerrung sollten die Schätzungen auf der 1. Winkelhalbierenden liegen.

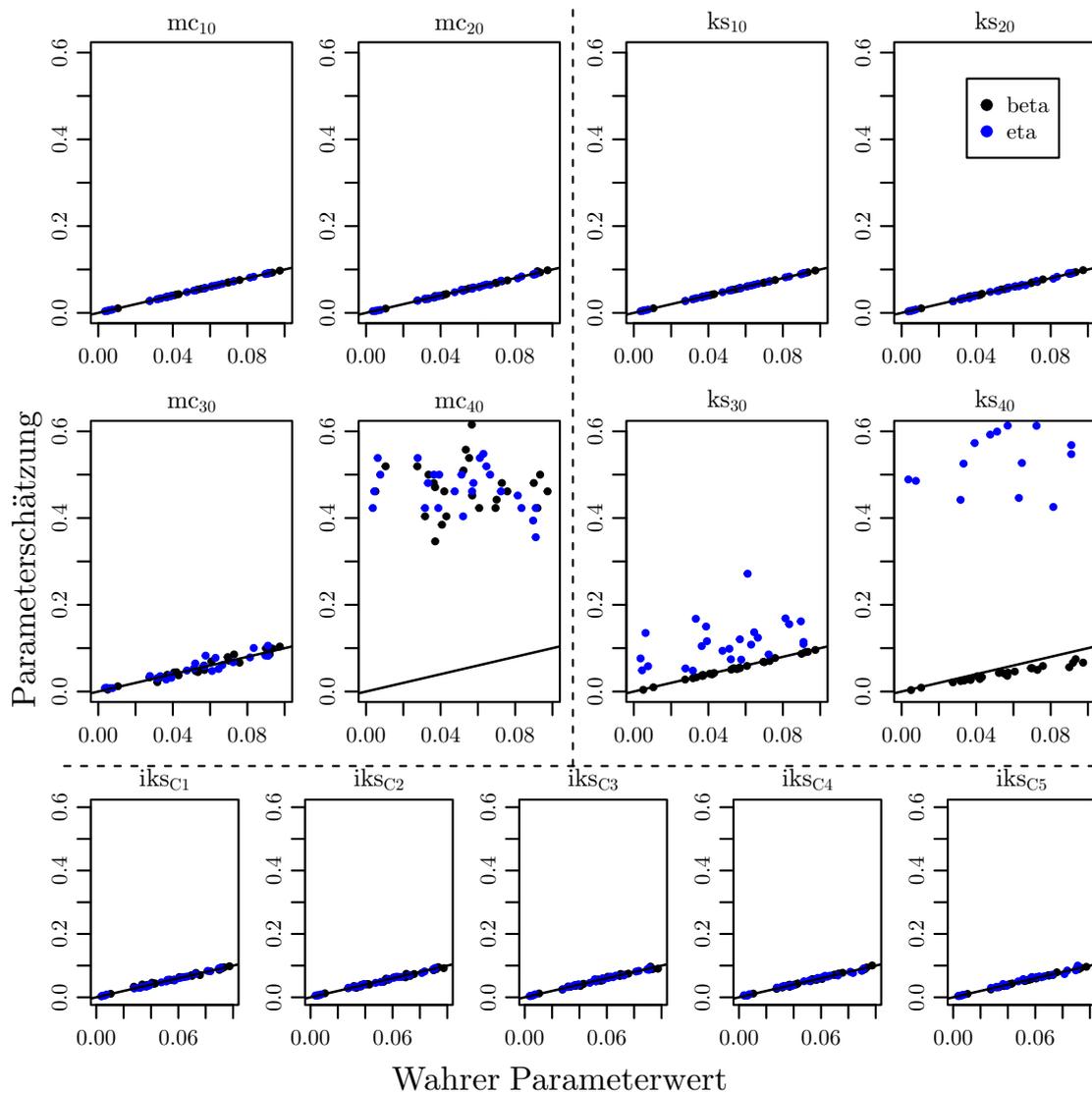


Abbildung C1. Parameterschätzung mit dem $BLIM_{COMP}$ in Abhängigkeit des wahren Parameterwertes. Die durchgehende Linie zeigt die 1. Winkelhalbierende.

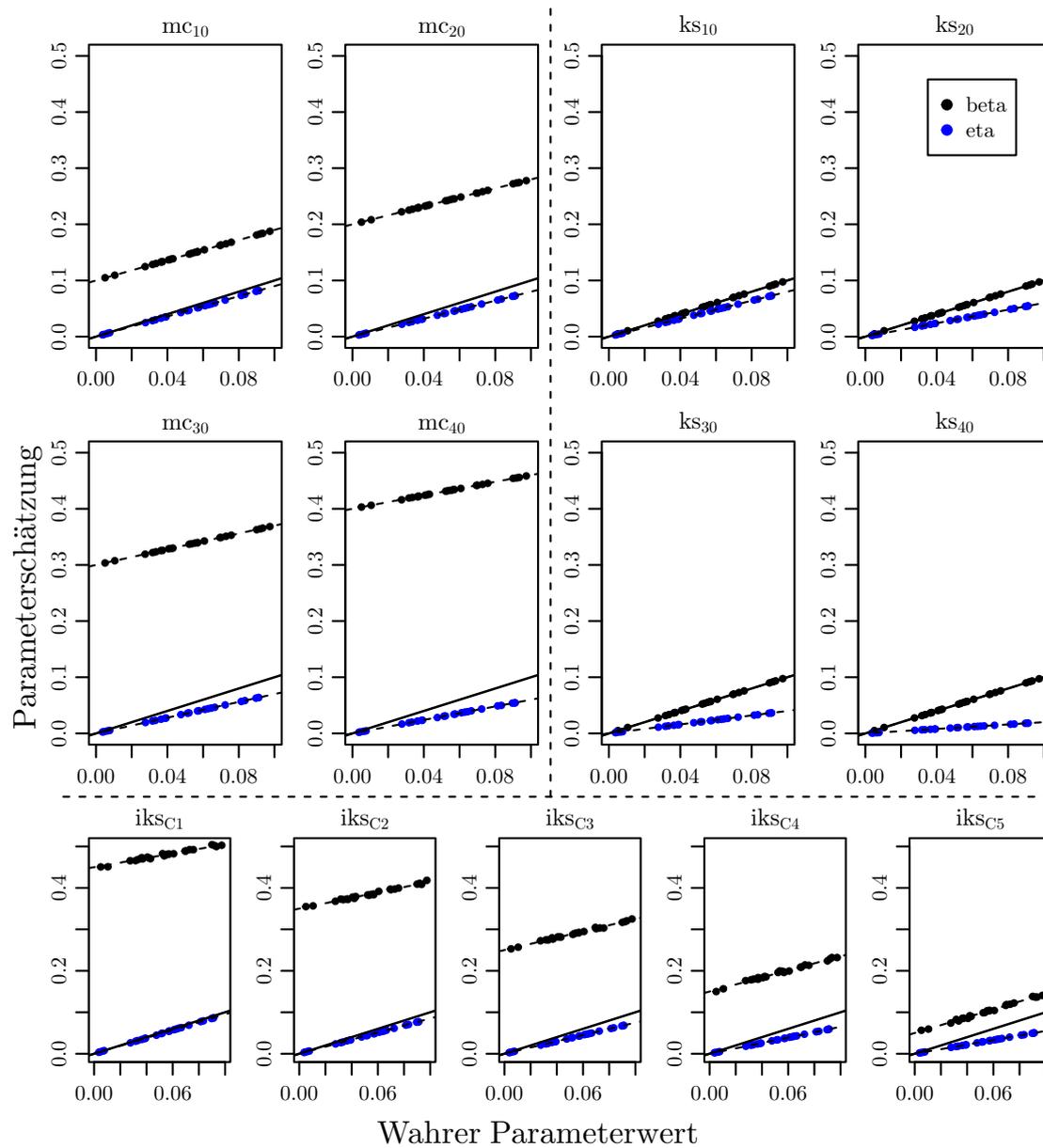


Abbildung C2. Parameterschätzung mit dem $BLIM_{MAW}$ in Abhängigkeit des wahren Parameterwertes. Die gestrichelte Linie zeigt die theoretisch zu erwartende Verzerrung, die durchgezogene Linie die 1. Winkelhalbierende.

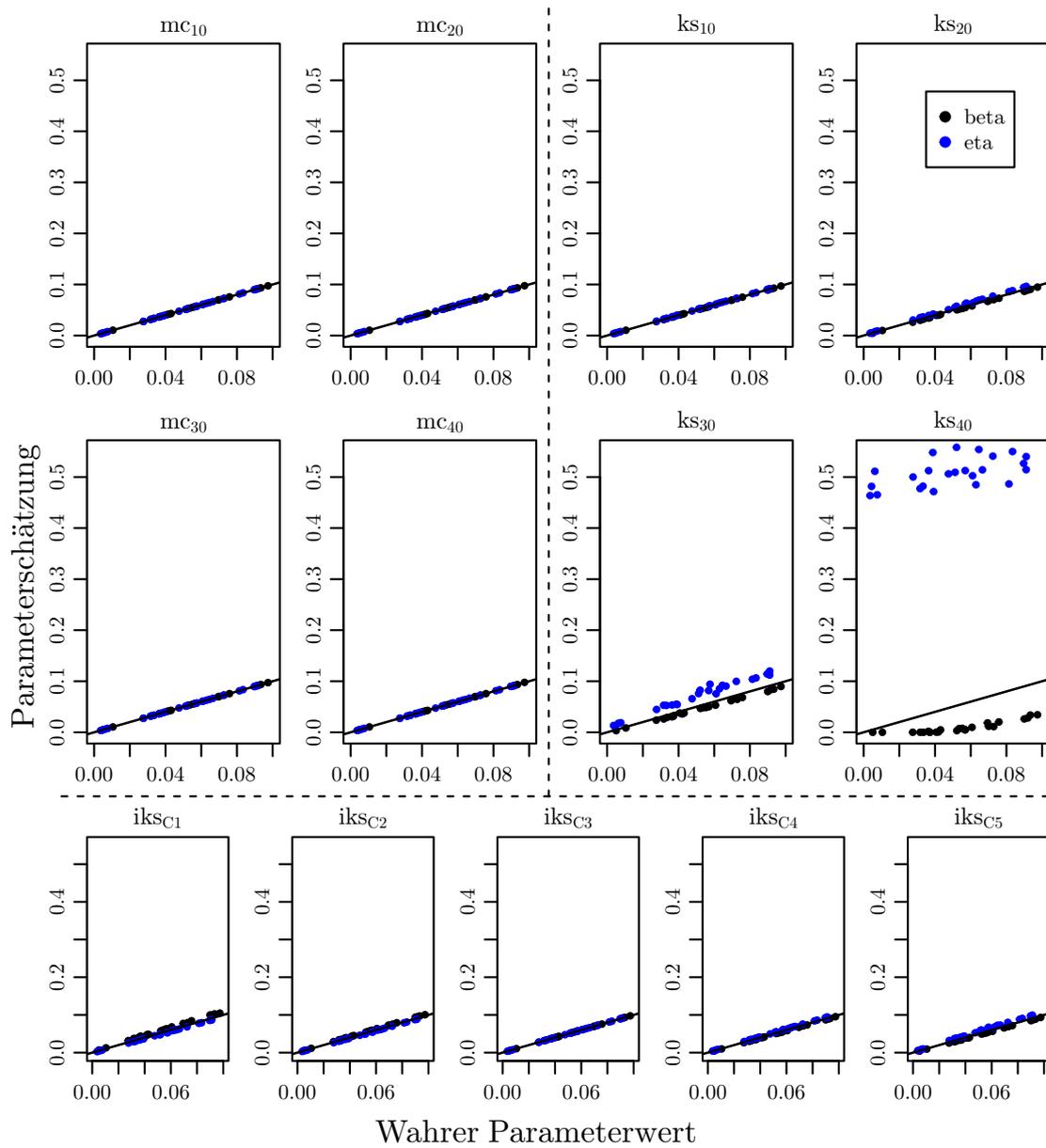


Abbildung C3. Parameterschätzung mit dem IMBLIM in Abhängigkeit des wahren Parameterwertes. Die durchgehende Linie zeigt die 1. Winkelhalbierende.

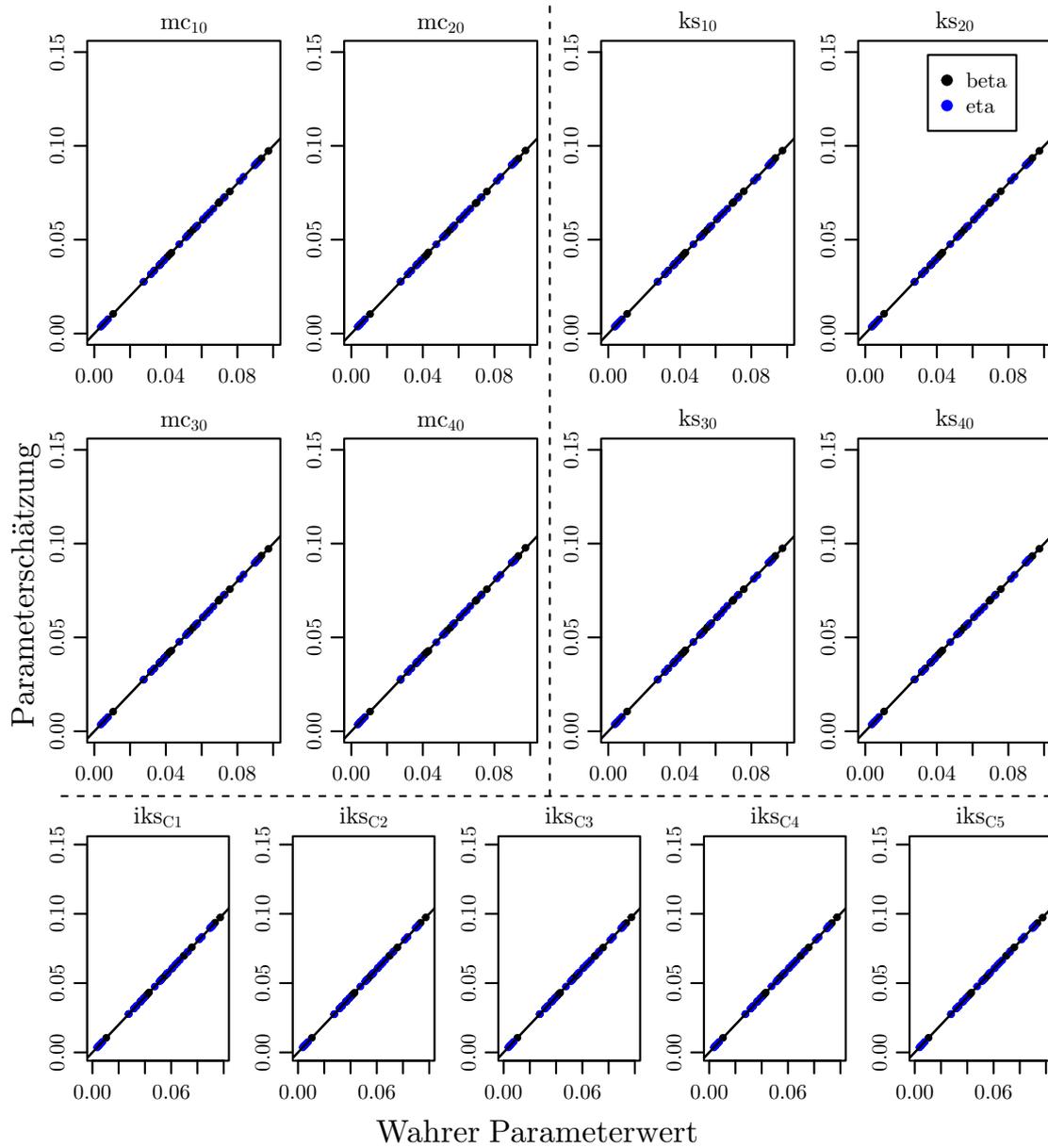


Abbildung C4. Parameterschätzung mit dem MissBLIM in Abhängigkeit des wahren Parameterwertes. Die durchgehende Linie zeigt die 1. Winkelhalbierende.

D η -Parameter und Stichprobengröße

Wiederherstellung des η -Parameters in Abhängigkeit der Stichprobengröße.

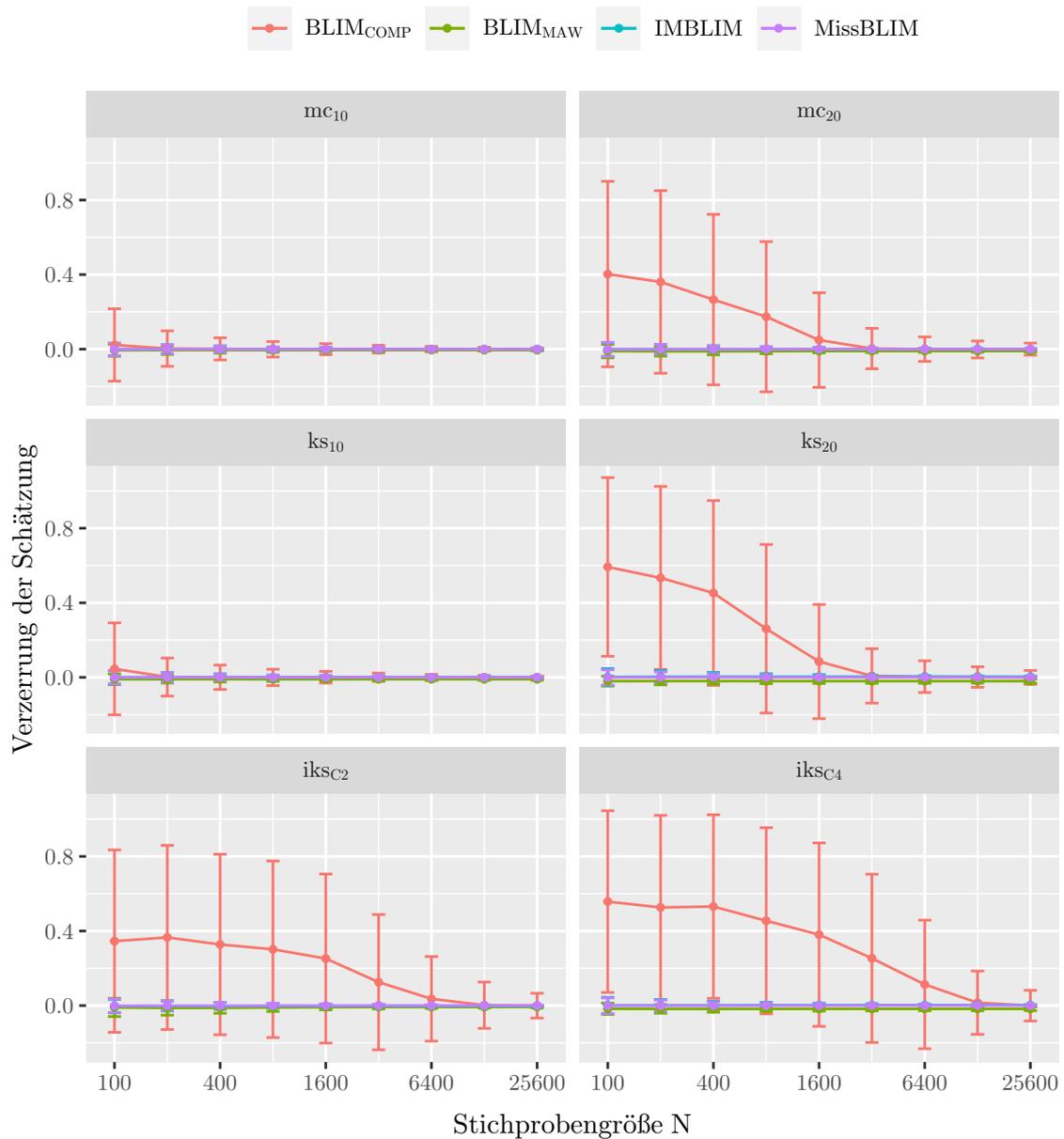


Abbildung D1. Verzerrung des η -Parameters der verschiedenen Modell gemittelt über alle Aufgaben in Abhängigkeit der simulierten Stichprobengröße N und des Dezimierungsmechanismus basierend auf der Wissensstruktur \mathcal{K}_s . Zusätzlich sind die Standardabweichungen als Fehlerbalken eingezeichnet. Die Abszisse ist zur Basis 2 logarithmiert.

E Tabellen empirische Wissensstrukturen

Nachfolgend die Tabellen der mittleren Verzerrung und Streuung für die drei empirischen Wissensstrukturen mit der Stichprobengröße $N = 3700$. Diese erleichtern den Vergleich mit den Ergebnissen basierend auf der simulierten Wissensstruktur.

Tabelle E1

Verzerrung und Standardabweichung der Parameterschätzungen in Abhängigkeit des Modells und Dezimierungsmechanismus über alle 12 Items gemittelt für die auf Basis der Wissensstruktur \mathcal{K}_{err} simulierten Daten.

	BLIM _{COMP}		BLIM _{MAW}		IMBLIM		MissBLIM	
	Bias β	Bias η	Bias β	Bias η	Bias β	Bias η	Bias β	Bias η
mc ₁	-16.72	-0.27	-15.42	-1.14	-16.26	-0.19	-16.30	-0.18
mc ₅	-18.56	-0.92	-13.57	-5.18	-16.94	-0.27	-17.00	-0.26
mc ₁₀	-19.46	-1.43	-10.23	-10.36	-16.84	-0.20	-16.94	-0.22
mc ₁₅	-19.30	-3.85	-6.91	-15.95	-17.57	-0.28	-17.78	-0.32
ks ₁	-16.32	-0.34	-16.03	-2.58	-16.04	-0.39	-15.39	-0.40
ks ₅	-16.58	-0.59	-15.11	-11.73	-15.57	-0.78	-11.77	-0.81
ks ₁₀	-16.14	-1.20	-13.14	-23.47	-14.57	-1.66	-7.10	-1.75
ks ₁₅	-16.33	-3.38	-10.80	-35.12	-13.04	-2.57	-4.69	-1.86
iks _{c1}	-20.74	-0.97	-9.55	0.00	-18.81	0.30	-13.05	-0.41
iks _{c2}	-19.23	-0.81	-12.14	-2.71	-17.56	-0.02	-16.81	-0.27
iks _{c3}	-18.06	-0.56	-13.85	-8.32	-15.95	-0.54	-16.18	-0.01
iks _{c4}	-16.30	-0.83	-14.08	-14.63	-15.06	-1.09	-12.04	-0.57
	BLIM _{COMP}		BLIM _{MAW}		IMBLIM		MissBLIM	
	SD β	SD η	SD β	SD η	SD β	SD η	SD β	SD η
mc ₁	32.36	7.51	31.72	7.81	31.85	7.25	31.92	7.25
mc ₅	34.73	9.37	33.51	13.34	32.10	7.32	32.19	7.33
mc ₁₀	37.72	11.94	37.35	22.37	32.68	7.31	32.74	7.31
mc ₁₅	46.16	20.92	41.40	32.85	33.31	7.39	33.50	7.42
ks ₁	31.25	7.69	30.73	8.32	31.11	7.01	29.96	6.81
ks ₅	31.67	12.36	29.09	22.74	31.66	7.65	24.70	6.77
ks ₁₀	32.26	25.24	25.73	44.65	31.68	9.12	17.65	7.49
ks ₁₅	33.42	57.42	21.76	66.44	31.46	11.28	13.71	7.41
iks _{c1}	38.11	6.90	38.35	10.39	33.07	6.46	25.23	6.00
iks _{c2}	35.49	7.91	35.34	11.12	32.20	6.68	31.76	6.72
iks _{c3}	33.90	10.74	31.34	18.03	31.37	7.57	30.83	7.18
iks _{c4}	31.77	14.65	28.15	28.60	31.12	8.32	24.82	7.15

Anmerkung. Alle Werte wurden mit 1000 multipliziert um die Darstellung zu verbessern. Die simulierte Stichprobengröße pro Datensatz war $N = 3700$ mit je 200 Datensätzen pro Simulationsart.

Tabelle E2

Verzerrung und Standardabweichung der Parameterschätzungen in Abhängigkeit des Modells und Dezimierungsmechanismus über alle 12 Items gemittelt für die auf Basis der Wissensstruktur \mathcal{K}_{geh} simulierten Daten.

	BLIM _{COMP}		BLIM _{MAW}		IMBLIM		MissBLIM	
	Bias β	Bias η	Bias β	Bias η	Bias β	Bias η	Bias β	Bias η
mc ₁	-1.20	-6.47	0.54	-6.33	-1.22	-6.02	-1.20	-6.05
mc ₅	-1.66	-9.09	7.60	-8.09	-1.51	-6.25	-1.52	-6.28
mc ₁₀	-2.39	-12.52	16.44	-10.24	-1.35	-6.46	-1.39	-6.50
mc ₁₅	-4.07	-15.28	24.60	-12.69	-1.80	-6.82	-1.80	-6.91
ks ₁	-1.39	-6.80	-1.46	-6.56	-1.34	-5.95	-1.35	-5.72
ks ₅	-1.30	-13.14	-1.59	-10.46	-0.87	-7.46	-1.27	-5.82
ks ₁₀	-2.08	-18.75	-1.39	-15.22	0.28	-9.85	-0.86	-5.70
ks ₁₅	-3.38	-24.49	-1.17	-19.96	1.34	-12.61	-0.49	-5.91
iks _{c1}	-3.28	-5.79	20.68	-6.26	-2.75	-4.77	-1.79	-4.67
iks _{c2}	-2.26	-7.45	11.70	-6.99	-2.09	-5.59	-1.70	-5.79
iks _{c3}	-1.51	-11.38	4.08	-9.26	-1.01	-7.00	-1.45	-6.26
iks _{c4}	-1.30	-14.37	-0.44	-11.66	-0.29	-8.17	-1.14	-5.74
	BLIM _{COMP}		BLIM _{MAW}		IMBLIM		MissBLIM	
	SD β	SD η	SD β	SD η	SD β	SD η	SD β	SD η
mc ₁	7.64	13.62	7.16	13.25	7.31	12.70	7.33	12.76
mc ₅	11.13	19.20	15.10	16.57	7.79	13.32	7.81	13.35
mc ₁₀	16.29	25.45	29.76	21.04	8.46	13.61	8.57	13.73
mc ₁₅	23.05	31.95	43.67	26.82	8.81	14.85	8.88	15.02
ks ₁	7.84	14.31	7.60	13.47	7.59	12.45	7.45	12.08
ks ₅	9.04	26.16	7.33	21.07	7.10	14.89	6.90	12.20
ks ₁₀	11.48	40.02	7.09	31.24	7.05	18.91	6.28	12.09
ks ₁₅	14.51	65.62	6.73	41.28	7.22	22.73	5.90	12.15
iks _{c1}	15.46	12.50	37.26	14.08	9.38	11.14	7.95	10.10
iks _{c2}	12.25	15.70	22.28	14.84	8.42	12.25	8.30	12.34
iks _{c3}	10.28	23.17	10.72	18.73	7.24	14.18	7.34	13.08
iks _{c4}	9.62	28.46	6.96	23.44	7.15	15.88	6.95	11.85

Anmerkung. Alle Werte wurden mit 1000 multipliziert um die Darstellung zu verbessern. Die simulierte Stichprobengröße pro Datensatz war $N = 3700$ mit je 200 Datensätzen pro Simulationsart.

Tabelle E3

Verzerrung und Standardabweichung der Parameterschätzungen in Abhängigkeit des Modells und Dezimierungsmechanismus über alle 12 Items gemittelt für die auf Basis der Wissensstruktur \mathcal{K}_{ges} simulierten Daten.

	BLIM _{COMP}		BLIM _{MAW}		IMBLIM		MissBLIM	
	Bias β	Bias η	Bias β	Bias η	Bias β	Bias η	Bias β	Bias η
mc ₁	-3.21	-11.74	-0.24	-11.84	-3.02	-11.14	-3.03	-11.26
mc ₅	-4.93	-14.70	10.90	-15.13	-3.28	-11.13	-3.29	-11.17
mc ₁₀	-7.56	-21.28	24.42	-20.32	-3.16	-12.12	-3.15	-12.17
mc ₁₅	-14.45	-30.94	38.19	-26.20	-3.45	-12.62	-3.50	-12.62
ks ₁	-3.03	-12.86	-3.07	-13.51	-2.94	-11.96	-2.86	-11.04
ks ₅	-3.08	-20.70	-2.80	-21.22	-2.00	-13.71	-2.14	-9.23
ks ₁₀	-3.57	-36.57	-2.47	-31.26	-0.74	-16.80	-1.25	-8.57
ks ₁₅	-5.30	-34.98	-2.26	-40.93	0.36	-20.29	-0.78	-7.64
iks _{c1}	-9.63	-10.88	30.43	-11.86	-4.91	-9.32	-3.35	-7.39
iks _{c2}	-6.06	-13.73	17.35	-13.83	-3.83	-11.03	-3.26	-11.26
iks _{c3}	-4.36	-17.47	5.06	-17.90	-3.02	-12.31	-3.29	-11.53
iks _{c4}	-2.79	-24.61	-1.16	-23.17	-1.49	-13.86	-2.16	-9.08
	BLIM _{COMP}		BLIM _{MAW}		IMBLIM		MissBLIM	
	SD β	SD η	SD β	SD η	SD β	SD η	SD β	SD η
mc ₁	11.87	28.12	10.57	26.86	11.07	27.16	11.12	27.44
mc ₅	17.00	31.97	17.38	27.16	11.53	27.57	11.54	27.56
mc ₁₀	26.69	41.98	32.27	29.78	12.23	28.55	12.26	28.69
mc ₁₅	48.00	63.57	48.14	34.23	12.85	29.12	12.91	29.04
ks ₁	11.30	29.80	11.08	27.97	11.13	28.28	10.67	26.13
ks ₅	11.67	37.60	10.19	30.02	10.02	28.63	8.74	20.10
ks ₁₀	14.31	64.76	10.16	38.58	9.90	30.70	8.22	17.63
ks ₁₅	15.34	76.75	9.89	49.61	9.69	32.75	7.61	15.65
iks _{c1}	27.82	27.83	39.90	29.01	12.99	26.91	11.00	20.95
iks _{c2}	19.34	31.04	25.09	27.80	12.11	27.87	11.60	27.12
iks _{c3}	14.95	34.98	13.52	28.15	11.37	28.29	11.26	28.18
iks _{c4}	12.95	43.36	10.11	31.24	10.20	28.37	9.52	22.49

Anmerkung. Alle Werte wurden mit 1000 multipliziert um die Darstellung zu verbessern. Die simulierte Stichprobengröße pro Datensatz war $N = 3700$ mit je 200 Datensätzen pro Simulationsart.

F Abbildungen empirische Wissensstrukturen

η -Schätzung: $N = 3700$

Nachfolgend sind für die empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} die η -Parameterschätzungen in Abhängigkeit der wahren Parameterwerte für eine Stichprobengröße von $N = 3700$ abgebildet. Dies geschieht für jedes Modell und jeden Dezimierungsmechanismus.

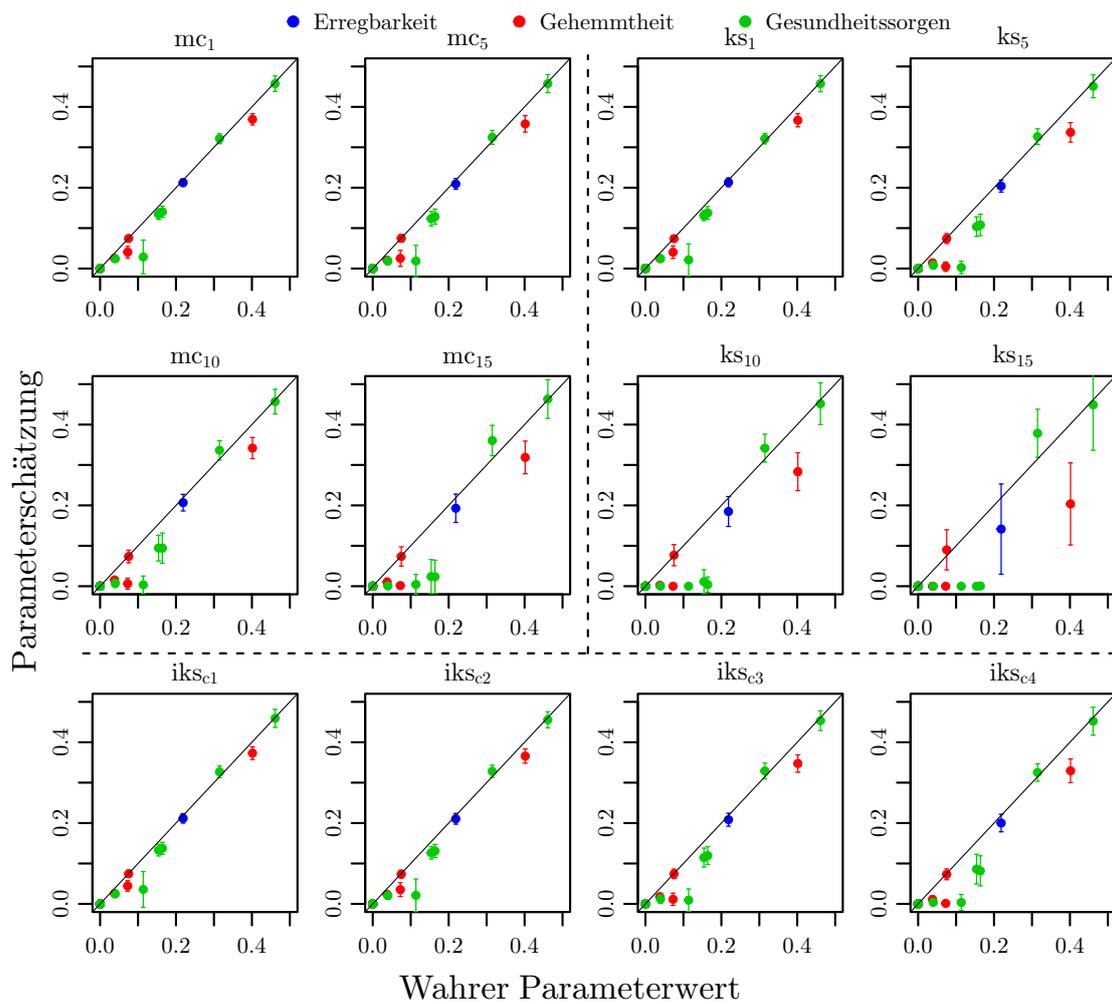


Abbildung F1. Mittlere η -Parameterschätzung des $BLIM_{COMP}$ in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 3700$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die Standardabweichung ist als Fehlerbalken eingezeichnet.

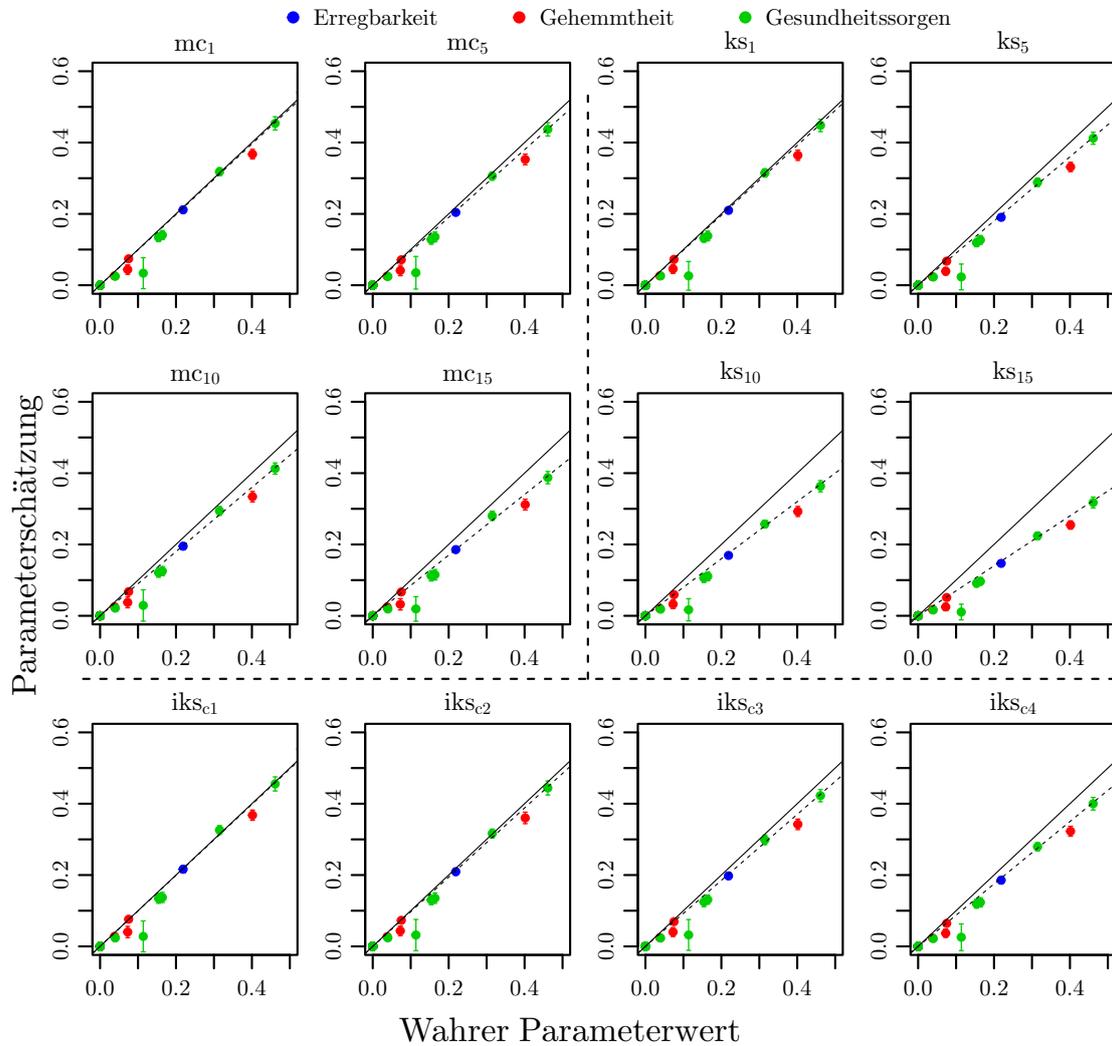


Abbildung F2. Mittlere η -Parameterschätzung des $BLIM_{MAW}$ in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 3700$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die gestrichelte die theoretisch zu erwartende Verzerrung. Die Standardabweichung ist als Fehlerbalken eingezeichnet.

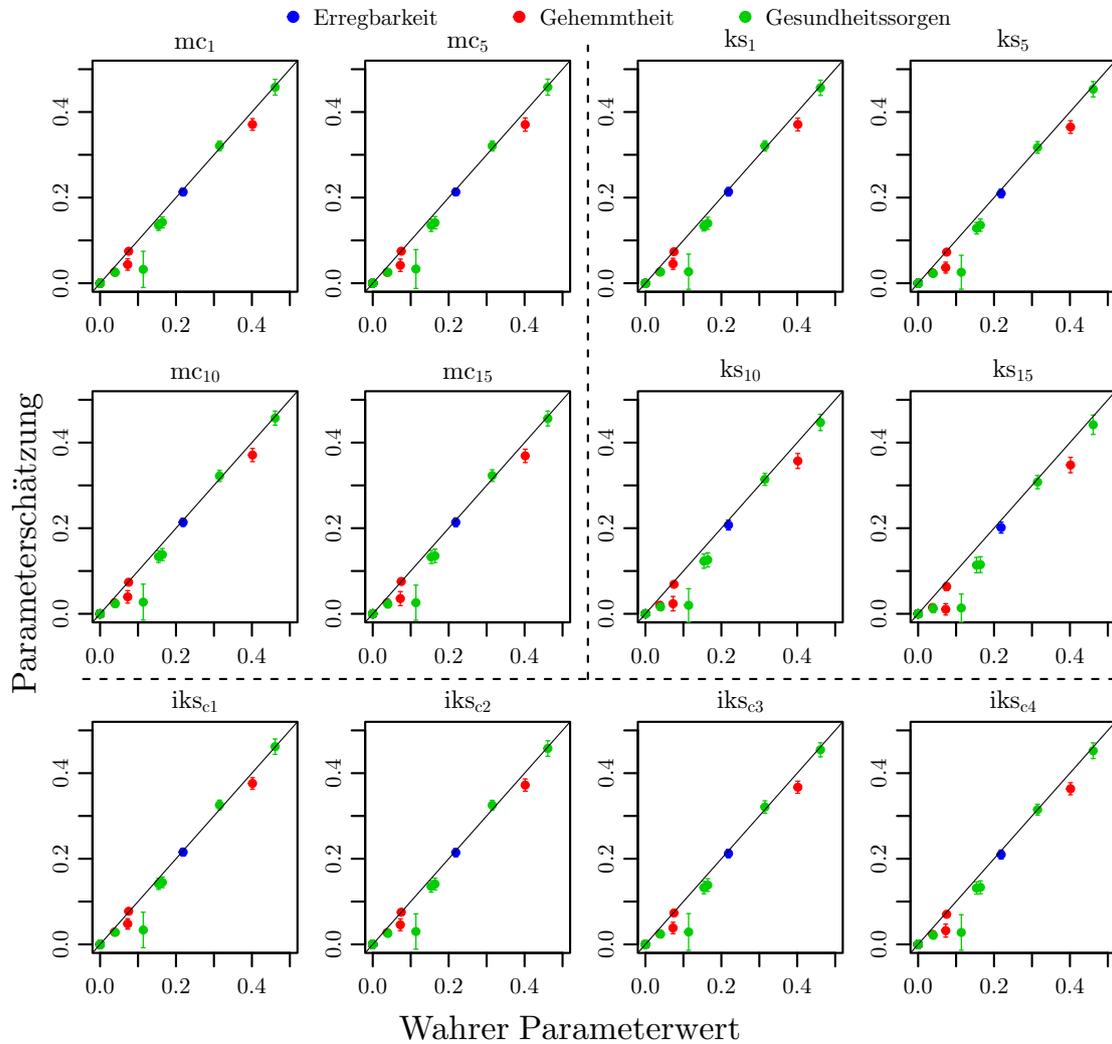


Abbildung F3. Mittlere η -Parameterschätzung des IMBLIM in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 3700$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die Standardabweichung ist als Fehlerbalken eingezeichnet.

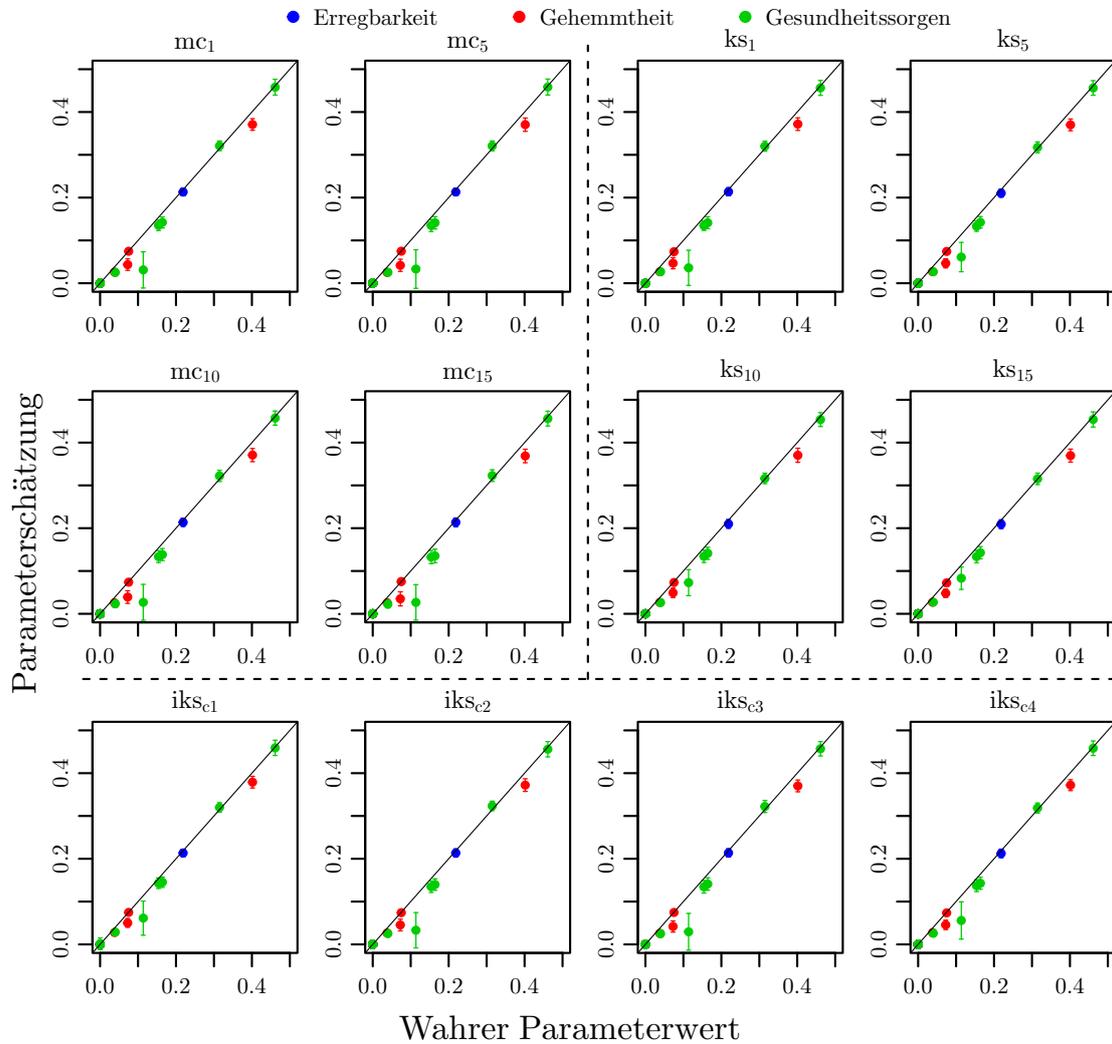


Abbildung F4. Mittlere η -Parameterschätzung des MissBLIM in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 3700$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die Standardabweichung ist als Fehlerbalken eingezeichnet.

β -Schätzung: $N = 10000$

Nachfolgend sind für die empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} die β -Parameterschätzungen in Abhängigkeit der wahren Parameterwerte für eine simulierte Stichprobengröße von $N = 10000$ abgebildet. Dies geschieht für jedes Modell und jeden Dezimierungsmechanismus. Bei keiner Verzerrung sollten die Schätzungen auf der 1. Winkelhalbierenden liegen.

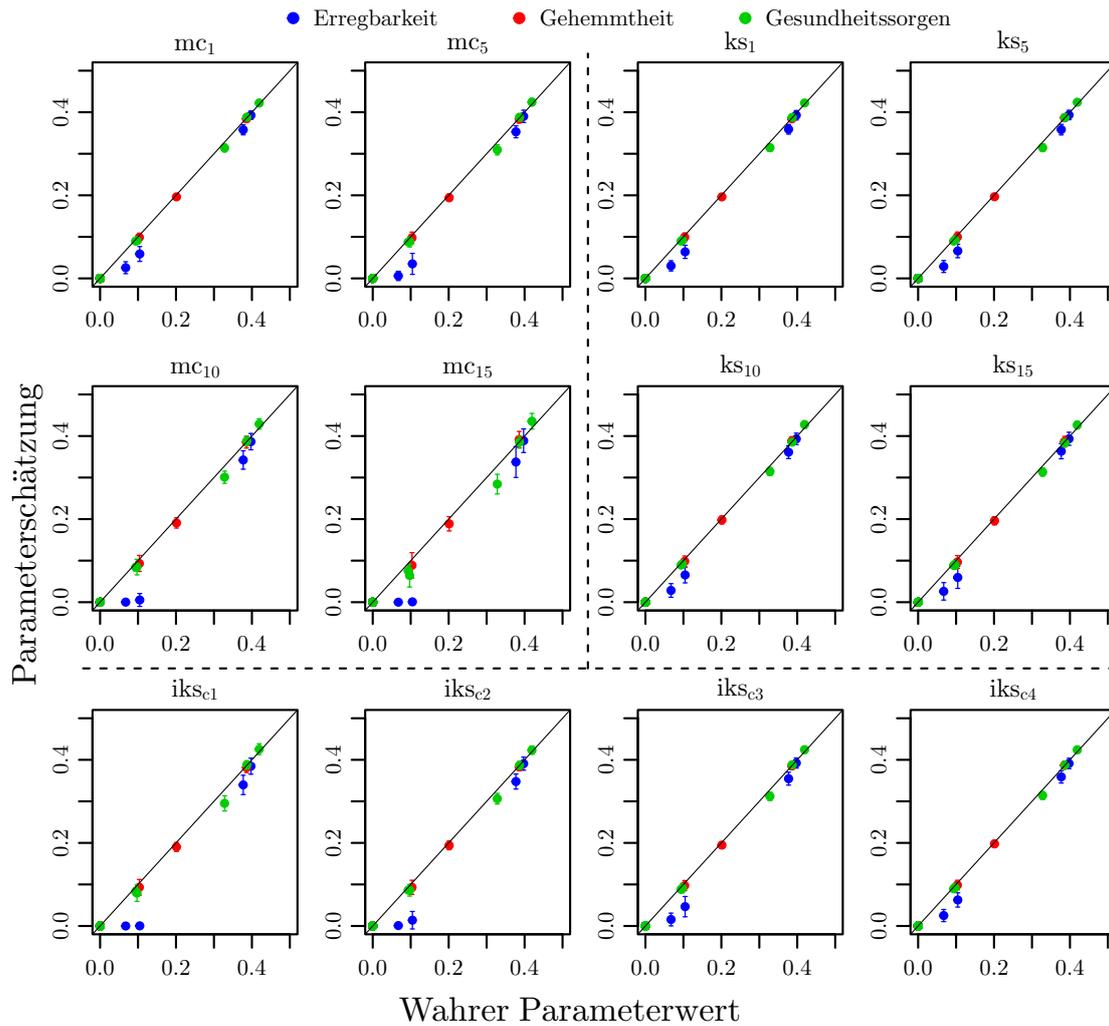


Abbildung F5. Mittlere β -Parameterschätzung des $BLIM_{COMP}$ in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 10000$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die Standardabweichung ist als Fehlerbalken eingezeichnet.

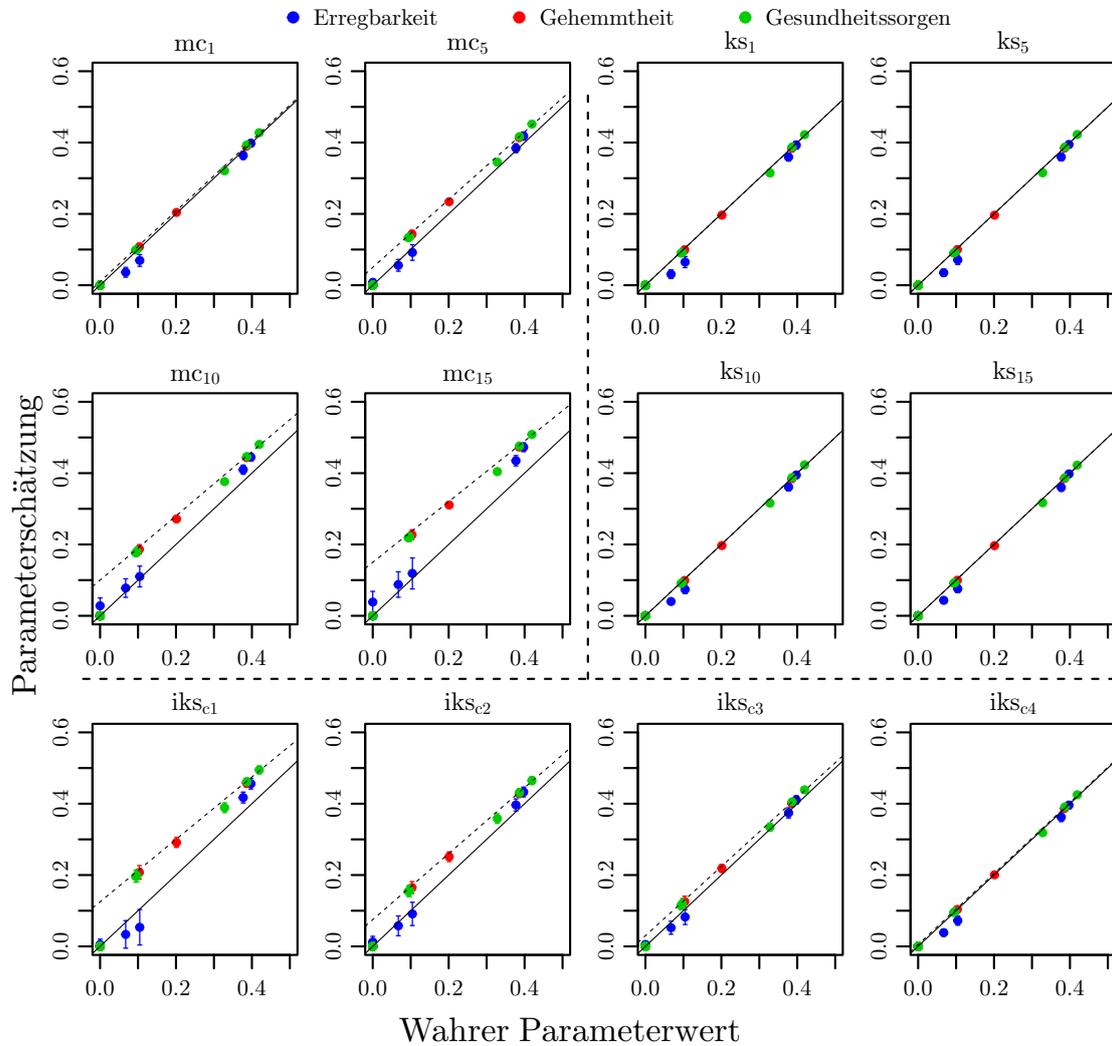


Abbildung F6. Mittlere β -Parameterschätzung des $BLIM_{MAW}$ in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 10000$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die gestrichelte die theoretisch zu erwartende Verzerrung. Die Standardabweichung ist als Fehlerbalken eingezeichnet.

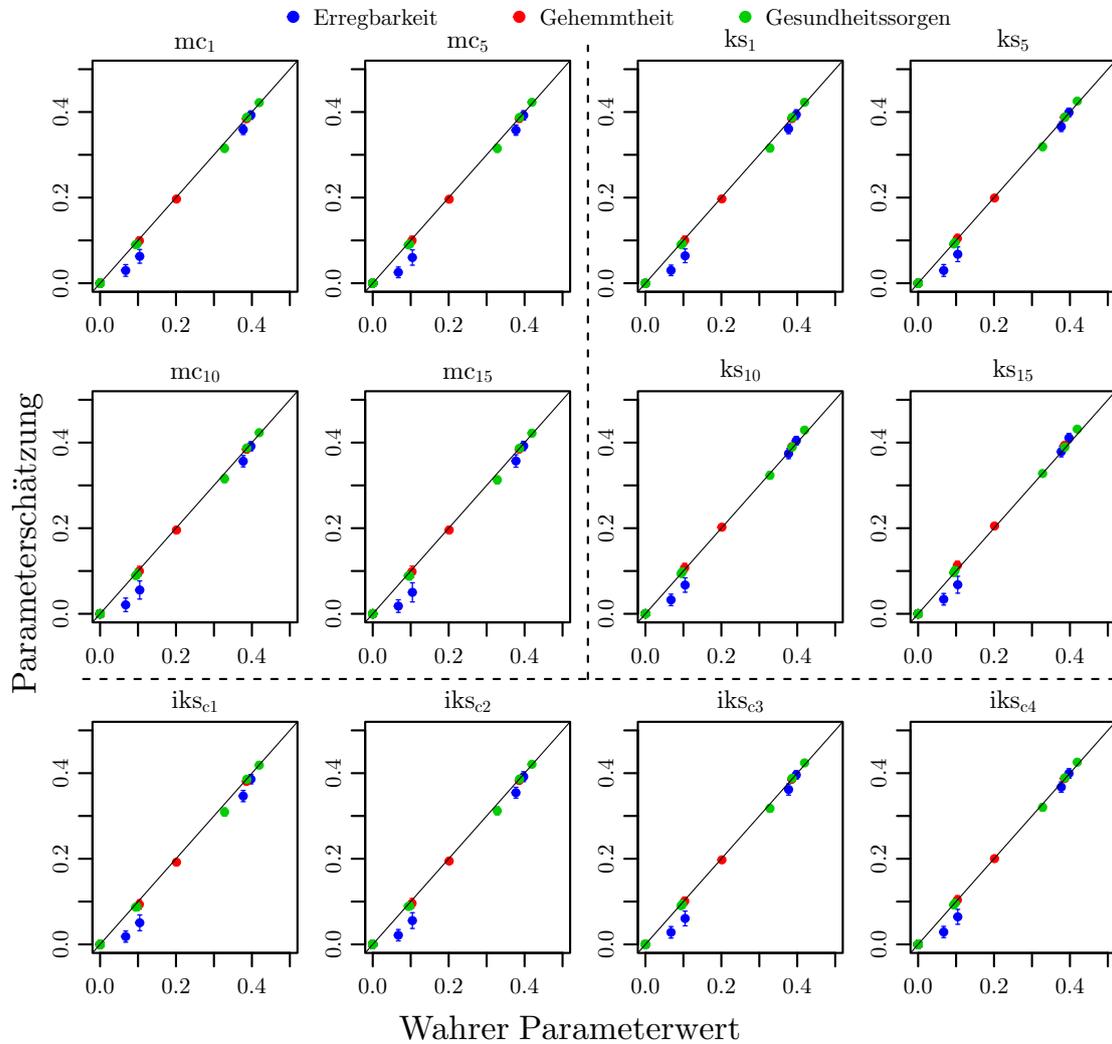


Abbildung F7. Mittlere β -Parameterschätzung des IMBLIM in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 10000$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die Standardabweichung ist als Fehlerbalken eingezeichnet.

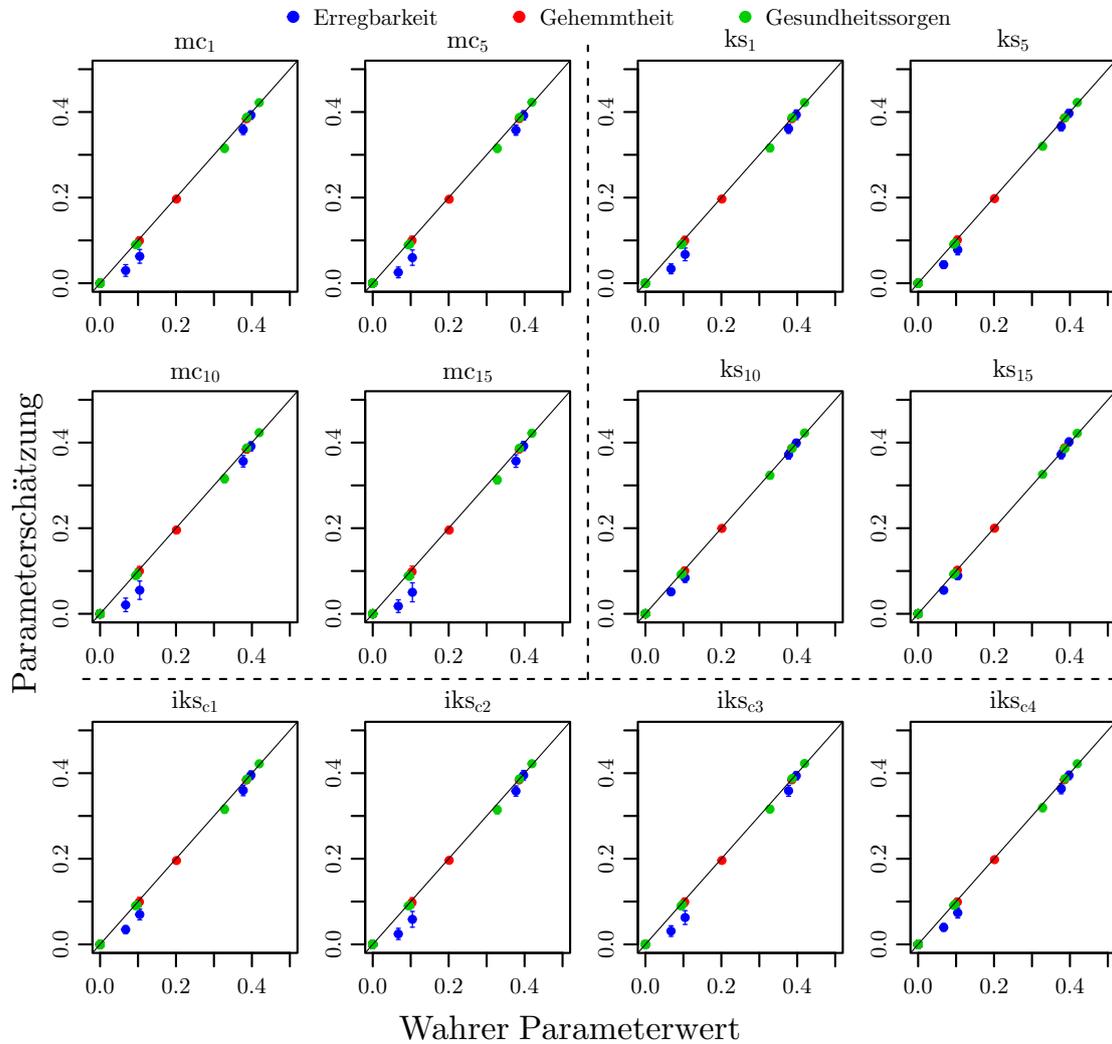


Abbildung F8. Mittlere β -Parameterschätzung des MissBLIM in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 10000$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die Standardabweichung ist als Fehlerbalken eingezeichnet.

η -Schätzung: $N = 10000$

Nachfolgend sind für die empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} die η -Parameterschätzungen in Abhängigkeit der wahren Parameterwerte für eine simulierte Stichprobengröße von $N = 10000$ abgebildet. Dies geschieht für jedes Modell und jeden Dezimierungsmechanismus. Bei keiner Verzerrung sollten die Schätzungen auf der 1. Winkelhalbierenden liegen.

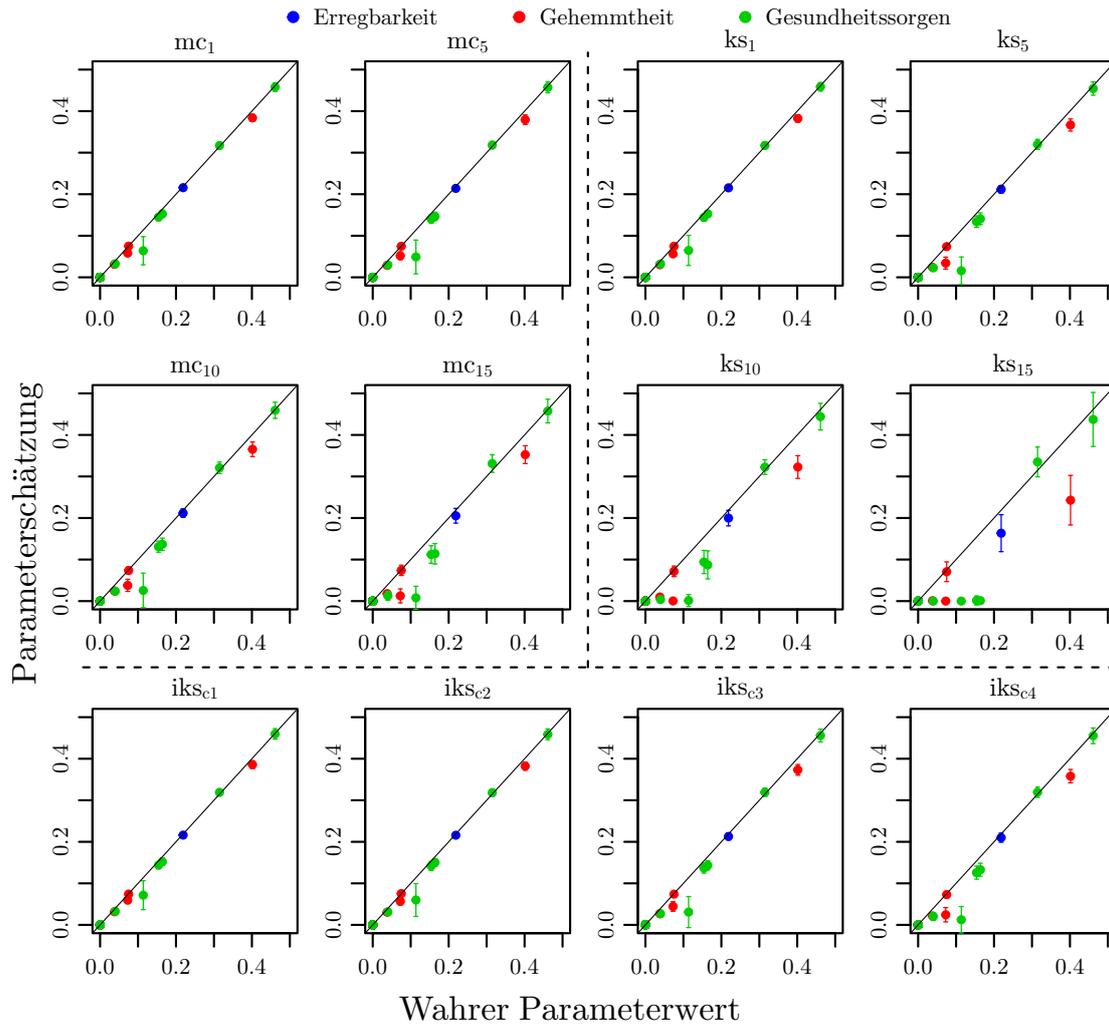


Abbildung F9. Mittlere η -Parameterschätzung des $BLIM_{COMP}$ in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 10000$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die Standardabweichung ist als Fehlerbalken eingezeichnet.

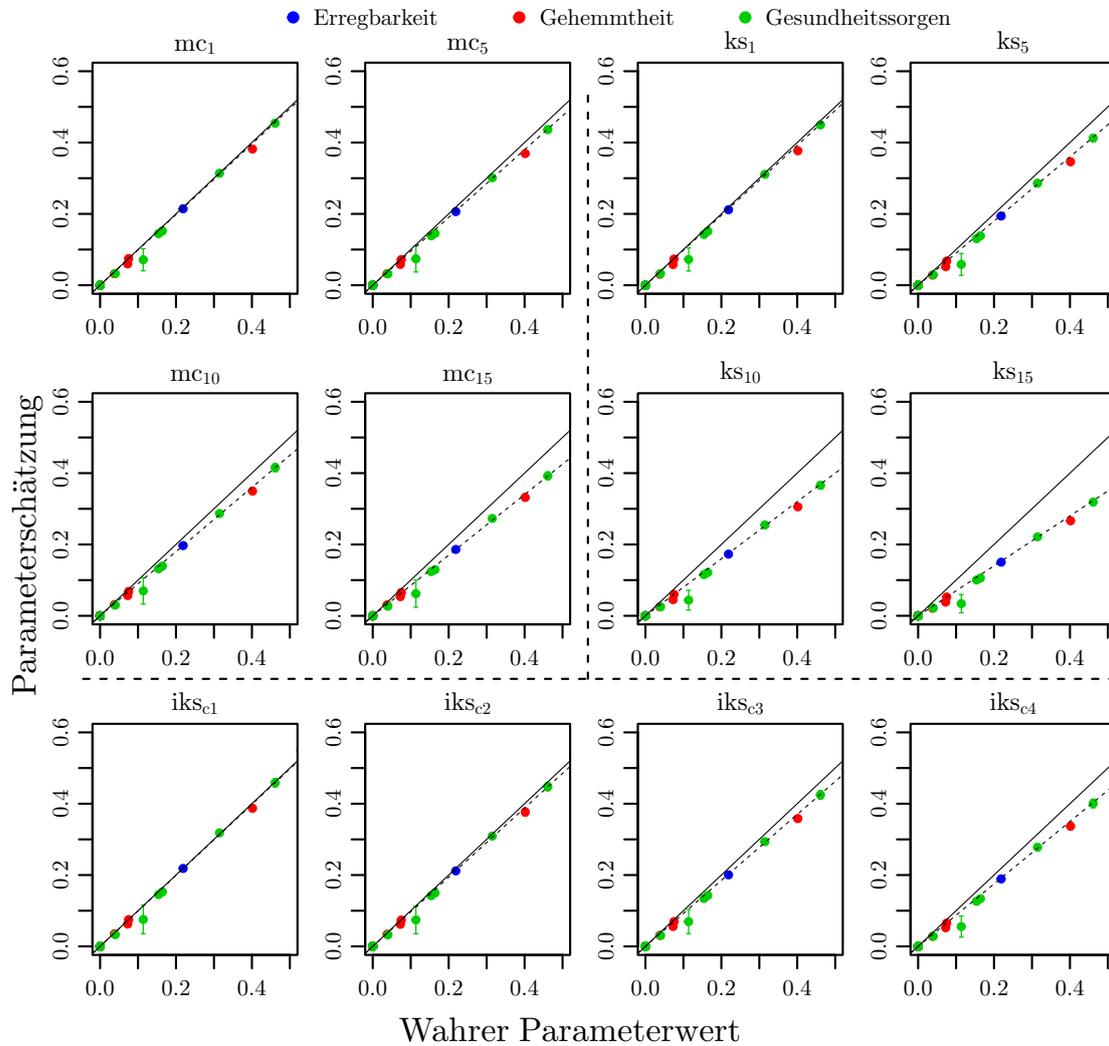


Abbildung F10. Mittlere η -Parameterschätzung des $BLIM_{MAW}$ in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 10000$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die gestrichelte die theoretisch zu erwartende Verzerrung. Die Standardabweichung ist als Fehlerbalken eingezeichnet.

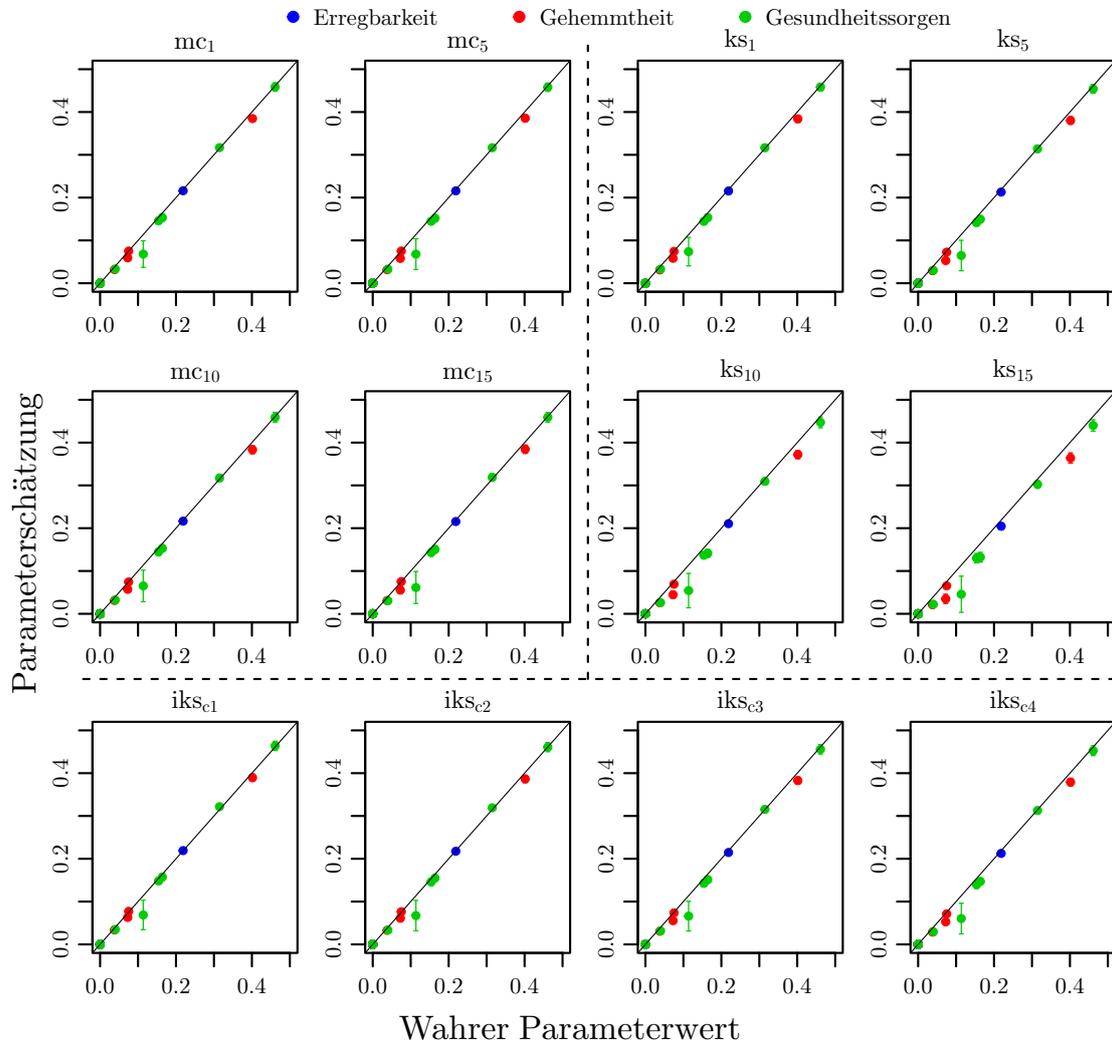


Abbildung F11. Mittlere η -Parameterschätzung des IMBLIM in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 10000$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die Standardabweichung ist als Fehlerbalken eingezeichnet.

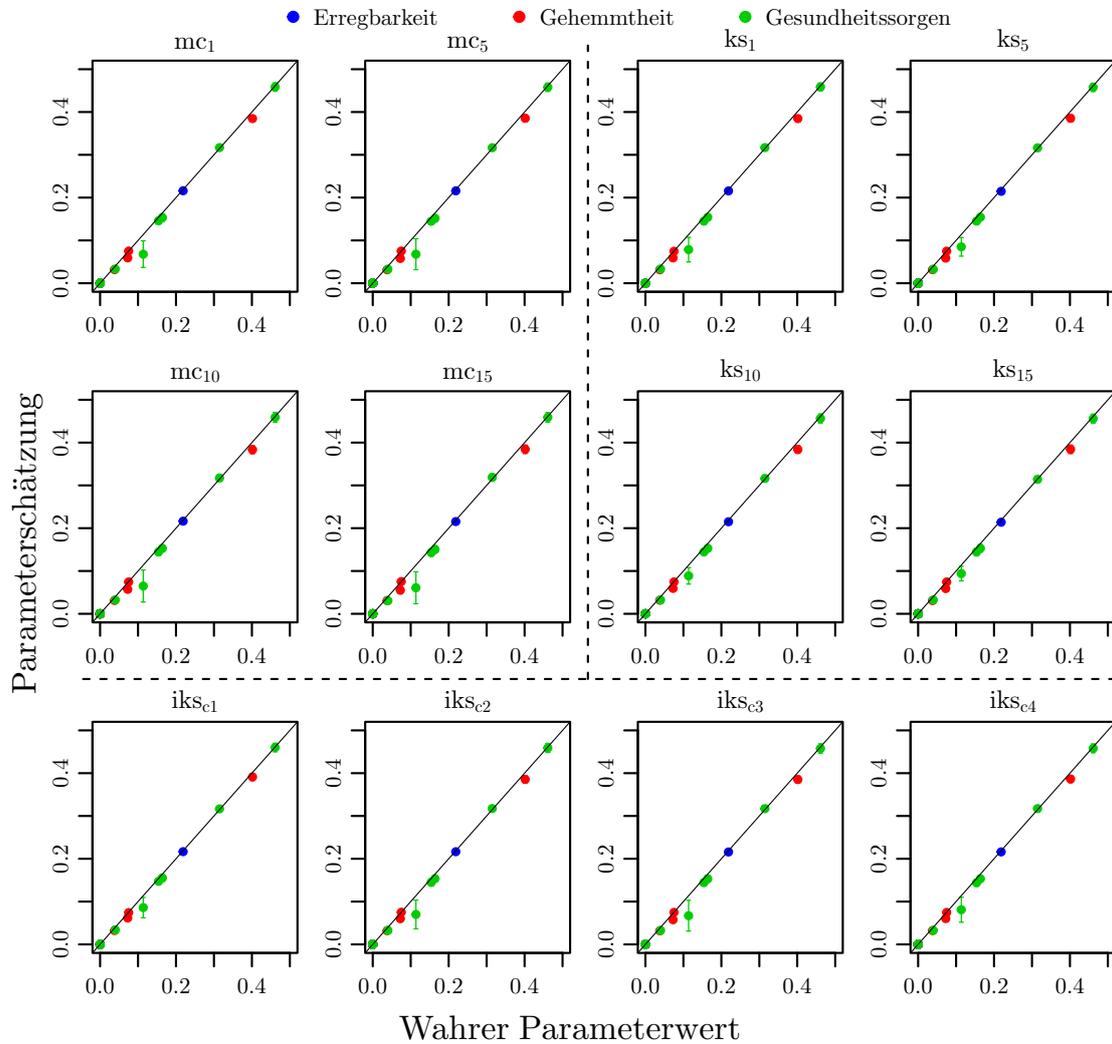


Abbildung F12. Mittlere η -Parameterschätzung des MissBLIM in Abhängigkeit des wahren Parameterwertes für jedes der 12 Items der drei empirischen Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine Stichprobengröße von $N = 10000$. Die durchgehende Linie zeigt die 1. Winkelhalbierende und die Standardabweichung ist als Fehlerbalken eingezeichnet.

Wiederherstellung des latenten Wissenszustands

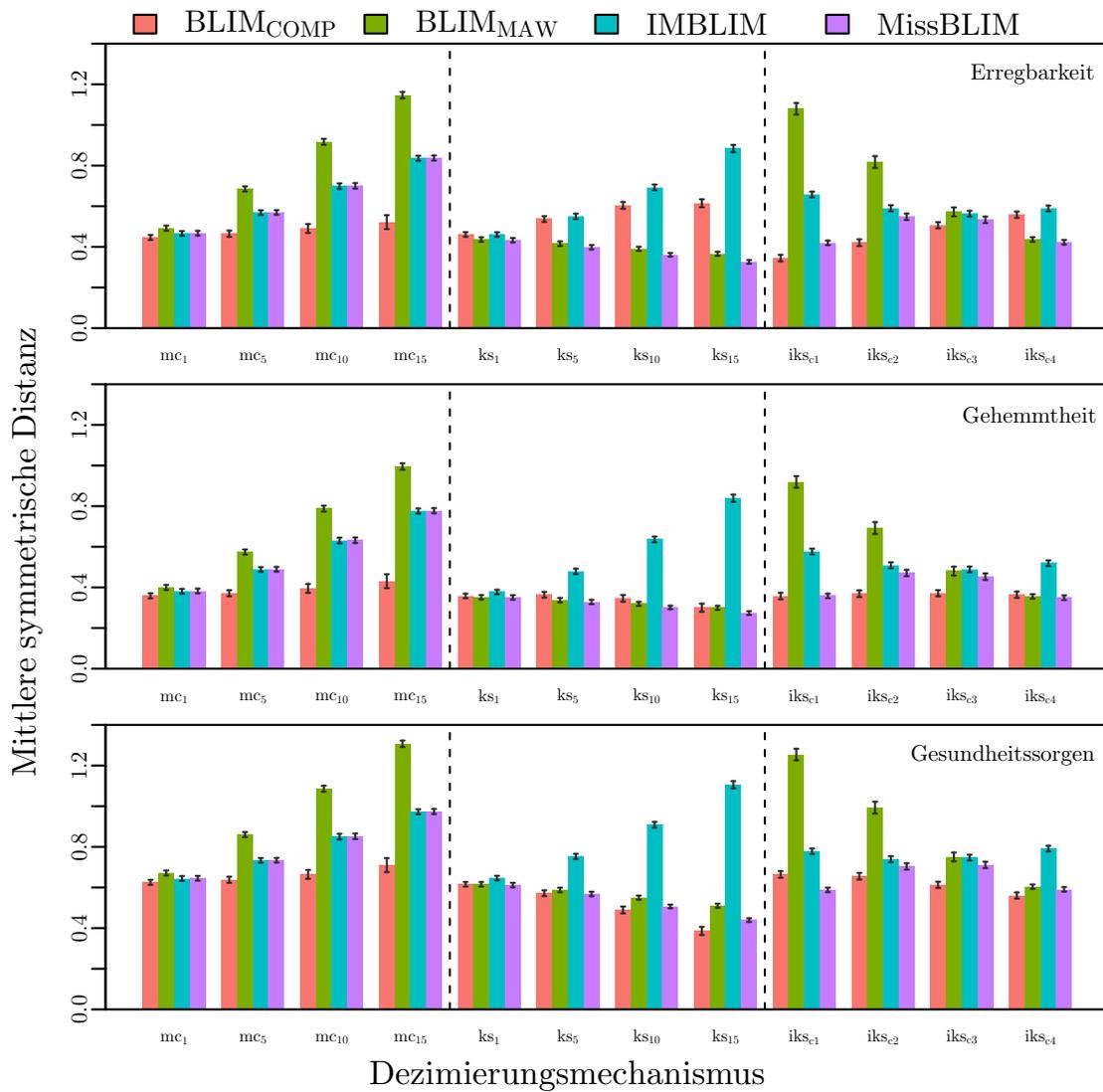


Abbildung F13. Mittlere symmetrische Distanz in Abhängigkeit des Modells und Dezimierungsmechanismus basierend auf den Skalen Erregbarkeit, Gehemmtheit und Gesundheitssorgen bzw. den unterliegenden Wissensstrukturen \mathcal{K}_{err} , \mathcal{K}_{geh} und \mathcal{K}_{ges} für eine simulierte Stichprobengröße mit $N = 10000$.