Behandlung von fehlenden Werten bei nicht ignorierbaren Ausfallmechanismen

Dissertation

zur Erlangung des akademischen Grades
doctor rerum politicarum
(Dr. rer. pol.)

vorgelegt dem Rat der Wirtschaftswissenschaftlichen Fakultät
der Friedrich-Schiller-Universität Jena

am 19.10.2004

von Diplom-Kaufmann Thomas Lehmann
geboren am 08.02.1974 in Gotha

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

ACMV	Available Case Missing Value
CCMV	Complete Case Missing Value
EM	Expectation Maximization
iid	unabhängig und identisch verteilt (engl.: independent and identically distributed)
MAR	Missing at Random
MCAR	Missing Completely at Random
ML	Maximum Likelihood
MNAR	Missing Not at Random
O.B.d.A.	Ohne Beschränkung der Allgemeinheit
PM	Pattern-Mixture
RPW	Response Propensity Weighting

Symbolverzeichnis

$\Gamma(\cdot)$ Gamma-Funktion

ε Parametervektor der Verteilung $P_\varepsilon(\underline{R})$

θ Parametervektor der Verteilung $P_\theta(\underline{Y})$

$\lambda, \lambda_1, \lambda_2$ Parameter zur Sensitivitätsbestimmung von Schätzungen

μ Erwartungswert(vektor)

$\mu_j^{(r)}$ Erwartungswert der Zufallsvariable $Y_j^{(r)}$

Σ Kovarianzmatrix

σ^2 Varianz

$\sigma_{jl}^{(r)}$ Kovarianz der Zufallsvariablen Y_j und Y_l im Pattern r (Pattern-Mixture Modell)

$\sigma_{22\cdot1}^{(r)}$ Varianz der Zufallsvariable $U^{(r)}$ in dem Regressionsmodell $Y_2^{(r)} = a^{(r)} + b^{(r)} Y_1^{(r)} + U^{(r)}$

τ Anteil der aus der Nichtbeobachtung resultierenden Varianz an der gesamten Varianz von Q (Multiple Imputation)

$\Phi(\cdot)$ Verteilungsfunktion der Standardnormalverteilung

$\phi(\cdot)$ Dichte der Standardnormalverteilung

ψ Parameter(vektor) der Verteilung $P_\psi(\mathbf{R} \mid \mathbf{y})$

$\mathbf{\Omega}_{\theta,\psi}$ Parameterraum von (θ, ψ)

ω Parametervektor der bedingten Verteilung $P_\omega(\underline{Y} \mid \underline{R})$

$\omega_{r,1\cdot2}$ Parametervektor der bedingten Verteilung $P_{\omega_{r,1\cdot2}}(Y_1 \mid Y_2 = y_2, R_2 = r)$

$\omega_{r,2\cdot1}$ Parametervektor der bedingten Verteilung $P_{\omega_{r,2\cdot1}}(Y_2 \mid Y_1 = y_1, R_2 = r)$

$a^{(r)}$ Parameter der Regression $Y_2^{(r)} = a^{(r)} + b^{(r)} Y_1^{(r)} + U^{(r)}$

B aus der multiplen Imputation resultierende Varianz des Schätzers (Between Imputation Variance)

$B(n,\varepsilon_1)$ Binomialverteilung mit den Parametern n und ε_1

$b^{(r)}$ Parameter der Regression $Y_2^{(r)} = a^{(r)} + b^{(r)}Y_1^{(r)} + U^{(r)}$

$Beta(\alpha,\beta)$ Betaverteilung mit den Parametern α und β

c relativer Anstieg der Varianz von Q aufgrund des Datenausfalls (Multiple Imputation)

$c^{(r)}$ Parameter der Regression $Y_1^{(r)} = c^{(r)} + d^{(r)}Y_2^{(r)} + V^{(r)}$

$\text{cov}(\cdot)$ Kovarianz

$d^{(r)}$ Parameter der Regression $Y_1^{(r)} = c^{(r)} + d^{(r)}Y_2^{(r)} + V^{(r)}$

df Anzahl der Freiheitsgrade

$E(\cdot)$ Erwartungswert

k Anzahl der betrachteten Merkmale in einer Untersuchung

$L(\cdot)$ Likelihood-Funktion

$l(\cdot)$ Loglikelihood-Funktion

M Zufallsvariable, welche die Anzahl der fehlenden Werte je Merkmalsträger in den einzelnen Pattern beschreibt

$M(n,\theta)$ Multinomialverteilung mit dem Parameter n und dem Parametervektor θ

m Anzahl der Ersetzungen je fehlendem Wert im Rahmen der Multiplen Imputation

N Anzahl der Merkmalsträger in der Grundgesamtheit

$N(\mu,\sigma^2)$ Normalverteilung mit Erwartungswert μ und Varianz σ^2

n Anzahl der Merkmalsträger in einer Stichprobe

n_{ab} Anzahl der Merkmalsträger mit $Y_1 = a$ und $Y_2 = b$ $(a,b = 0,1)$

n_{ab}^R absolute Häufigkeit der vollständig beobachteten Merkmalsträger mit $Y_1 = a$ und $Y_2 = b$

n^{NR}	Anzahl der nicht beobachteten Merkmalsträger in einer Stichprobe
p_{ab}	Wahrscheinlichkeit, dass die Zufallsvariablen Y_1 und Y_2 die Werte a bzw. b annehmen $(a,b = 0,1)$
q	Anzahl der vollständig beobachteten Merkmalsträger in einer Stichprobe
Q	Kennzahl in der Grundgesamtheit
$P_\theta(\underline{Y})$	durch θ beschriebene Verteilung der Zufallsvariable $\underline{Y}$
$P_\theta(\mathbf{y})$	Wahrscheinlichkeit bzw. Dichte der Datenmatrix $\mathbf{y}$
R_j	Zufallsvariable, welche die (Nicht-)Beobachtung von Y_j beschreibt ($j = 1,\ldots,k$)
R_{ij}	Zufallsvariable, welche die (Nicht-)Beobachtung von Y_{ij} beschreibt ($i = 1,\ldots,n; j = 1,\ldots,k$)
$R_i^{(1)}$	Unit Nonresponse Indikatorvariable für Untersuchungseinheit i
$R_{ij}^{(2)}$	Item Nonresponse Indikatorvariable für das Merkmal j in Untersuchungseinheit i ($R_{ij}^{(2)} = R_{ij}$)
$\underline{R}$	(mehrdimensionale) Zufallsvariable mit $\underline{R} = (R_1,...,R_k)$
$\underline{R}_i$	(mehrdimensionale) Zufallsvariable mit $\underline{R}_i = (R_{i1},\ldots,R_{ik})$ $(i = 1,..,n)$
$\mathbf{R}$	Indikatormatrix mit $\mathbf{R} = (\underline{R}_1,\ldots,\underline{R}_n)^T$
$\mathbf{R}^{(1)}$	Unit Nonresponse Indikatormatrix mit $\mathbf{R}^{(1)} = (R_1^{(1)},\ldots,R_n^{(1)})^T$
$\mathbf{R}^{(2)}$	Item Nonresponse Indikatormatrix ($\mathbf{R}^{(2)} = \mathbf{R}$)
r	Pattern im Pattern-Mixture Modell
r_j	Realisation der Zufallsvariable R_j ($j = 1,\ldots,k$)
r_{ij}	Realisation der Zufallsvariable R_{ij} ($i = 1,..,n; j = 1,\ldots,k$)
$\underline{r}$	Realisation der Zufallsvariable $\underline{R}$ mit $\underline{r} = (r_1,...,r_k)$
$\underline{r}_i$	Realisation der Zufallsvariable $\underline{R}_i$ mit $\underline{r}_i = (r_{i1},\ldots,r_{ik})$ $(i = 1,..,n)$
$\mathbf{r}$	Realisation der Indikatormatrix $\mathbf{R}$ mit $\mathbf{r} = (\underline{r}_1,\ldots,\underline{r}_n)^T$
s^2	Stichprobenvarianz

s_R^2 Varianz der beobachteten q Werte in einer Stichprobe

$t^{df}_{1-\frac{\alpha}{2}}$ $\left(1-\frac{\alpha}{2}\right)$-Fraktil der t-Verteilung mit df Freiheitsgraden

T Varianz des Schätzers für Q im Rahmen der multiplen Imputation

t Iteration des EM-Algorithmus bzw. des Data-Augmentation Verfahrens

U Varianz eines Stichprobenschätzers für Q bei einer einzelnen Ersetzung im Rahmen der multiplen Imputation

$U^{(r)}$ Residualvariable der Regression $Y_2^{(r)} = a^{(r)} + b^{(r)}Y_1^{(r)} + U^{(r)}$

$\overline{U}$ durchschnittliche Varianz eines Stichprobenschätzers für Q bei einer einzelnen Ersetzung (Average Within-Imputation Variance)

$V^{(r)}$ Residualvariable der Regression $Y_1^{(r)} = c^{(r)} + d^{(r)}Y_2^{(r)} + V^{(r)}$

Var(·) Varianz

Y_j Zufallsvariable, welche das j-te Merkmal beschreibt ($j = 1,\ldots,k$)

$Y_j^{(r)}$ Zufallsvariable im Pattern r ($Y_j^{(r)} := Y_j \mid R_2 = r$)

Y_{ij} Zufallsvariable, welche das j-te Merkmal des i-ten Merkmalsträgers beschreibt ($i = 1,\ldots,n$; $j = 1,\ldots,k$)

$\underline{Y}$ (mehrdimensionale) Zufallsvariable mit $\underline{Y} = (Y_1,...,Y_k)$

$\underline{Y}_i$ (mehrdimensionale) Zufallsvariable mit $\underline{Y}_i = (Y_{i1},\ldots,Y_{ik})$ ($i = 1,..,n$)

$\mathbf{Y}$ Zufallsmatrix mit $\mathbf{Y} = (\underline{Y}_1,\ldots,\underline{Y}_n)^T$

y_j Realisation der Zufallsvariable Y_j ($j = 1,\ldots,k$)

y_{ij} Realisation der Zufallsvariable Y_{ij} ($i = 1,..,n$; $j = 1,\ldots,k$)

$\overline{Y}_j$ Mittelwert der Zufallsvariable Y_j in der Grundgesamtheit

$\overline{y}_j$ Mittelwert der Realisationen von der Zufallsvariable Y_j in einer Stichprobe

$\overline{y}_j^R$ Mittelwert der beobachteten Realisationen von der Zufallsvariable Y_j in einer Stichprobe

$\bar{y}_{j,l}^{R}$ Mittelwert der beobachteten Werte der Zufallsvariable Y_j in Klasse l (Gewichtungsverfahren)

$\bar{y}_{j}^{RPW}$ Schätzer für den Mittelwert der Zufallsvariable Y_j innerhalb des Gewichtungsverfahrens

$\underline{y}$ Realisation der Zufallsvariable $\underline{Y}$ mit $\underline{y} = (y_1,...,y_k)$

$\underline{y}_i$ Realisation der Zufallsvariable $\underline{Y}_i$ mit $\underline{y}_i = (y_{i1},\ldots,y_{ik})$ $(i = 1,..,n)$

$\mathbf{y}$ Datenmatrix mit $\mathbf{y} = (\underline{y}_1,\ldots,\underline{y}_n)^T$

$\mathbf{Y}_{mis}$ nicht beobachtete Zufallsvariablen der Zufallsmatrix $\mathbf{Y}$

$\mathbf{Y}_{obs}$ beobachtete Zufallsvariablen der Zufallsmatrix $\mathbf{Y}$

$\underline{y}_{mis,i}$ nicht beobachtete Werte in Untersuchungseinheit i $(i = 1,..,n)$

$\underline{y}_{obs,i}$ beobachtete Werte in Untersuchungseinheit i $(i = 1,..,n)$

$\mathbf{y}_{mis}$ unbeobachteter Teil von $\mathbf{y}$ mit $\mathbf{y}_{mis} = (\underline{y}_{mis,1},\ldots, \underline{y}_{mis,n})^T$

$\mathbf{y}_{obs}$ beobachteter Teil von $\mathbf{y}$ mit $\mathbf{y}_{obs} = (\underline{y}_{obs,1},\ldots, \underline{y}_{obs,n})^T$

$z_{1-\frac{\alpha}{2}}$ $\left(1-\frac{\alpha}{2}\right)$-Fraktil der Standardnormalverteilung

$\propto$ ist proportional zu

$\sim$ ist verteilt wie

$\perp\!\!\!\perp$ unabhängig

∞ unendlich

1 Einleitung

1.1 Darstellung der Problematik

Fehlende Werte stellen ein häufig anzutreffendes Problem im Rahmen von empirischen Untersuchungen dar. Die Gründe für die unvollständige Beobachtung von Daten können dabei sehr vielfältig sein. Werden die Daten beispielsweise durch eine Befragung erhoben, kann der Ausfall u.a. durch Antwortverweigerung, mangelndes Wissen des Befragten oder durch schlichtes Übersehen einzelner Fragen verursacht werden.[1] Treten aus den verschiedensten Gründen fehlende Werte in einer Untersuchung auf, können die auf einem vollständigen Datenmaterial basierenden, statistischen Auswertungsmethoden nicht mehr unmittelbar zur Anwendung kommen. Durch eine Vernachlässigung der unvollständigen Datensätze ist eine Datenanalyse mittels statistischer Standardverfahren zwar prinzipiell durchführbar, jedoch ist diese Vorgehensweise nur gerechtfertigt, wenn kein Zusammenhang zwischen den erhobenen Merkmalen und der Nichtbeobachtung besteht. Ist diese Voraussetzung nicht erfüllt, sind die vollständigen Datensätze nicht repräsentativ für die Grundgesamtheit und die Datenqualität ist durch den Ausfall erheblich beeinträchtigt.[2] Da in der statistischen Literatur die einhellige Meinung besteht, dass in nahezu allen empirischen Untersuchungen nicht von der Gültigkeit dieser Annahme ausgegangen werden kann, ist eine Beschränkung der Analyse auf die beobachteten Datensätze im Allgemeinen nicht gerechtfertigt.[3]

Vor diesem Hintergrund sind in der Vergangenheit zahlreiche Verfahren entwickelt worden, die unter weniger restriktiven Voraussetzungen eine geeignete Behandlung von fehlenden Werten erlauben. Die Methoden basieren dabei fast ausnahmslos auf der Annahme, dass der Datenausfall durch die beobachteten Merkmalswerte erklärt werden kann. Obwohl dieser Zusammenhang in empirischen Untersuchungen aufgrund der unbekannten fehlenden Werte nicht überprüfbar ist, erfolgt in einer Vielzahl von publizierten Studien eine Behandlung der nicht beobachteten Daten durch

---

[1] Vgl. Bankhofer/Praxmarer (1998), S. 109.
[2] Vgl. Rässler (2000), S. 67; Schafer (1997), S. 1f.
[3] Vgl. Schafer/Olsen (1998), S. 11; Huisman (1998); Schnell (2002); Lillard et al. (1986); Landerman et al. (1997), S. 6; Crawford et al. (1995).

diese Verfahren. In vielen Fällen sind jedoch Zweifel an der Plausibilität dieser Annahme berechtigt, wie beispielsweise die Auswertungen von Einkommenserhebungen im Rahmen von Befragungen zeigen.[4] Diese Problematik wird durch die tendenziell sinkende Antwortbereitschaft in Bevölkerungsumfragen noch verschärft, so dass der adäquaten Behandlung von fehlenden Werten eine steigende Bedeutung beizumessen ist.[5]

Unter diesem Gesichtspunkt stehen in zunehmenden Maße Verfahren im Mittelpunkt, die unverzerrte Analyseergebnisse auch in Fällen ermöglichen, in denen der Datenausfall nicht durch die beobachteten Werte erklärt werden kann. Der zum Ausfall führende Mechanismus wird dabei als nicht ignorierbar bezeichnet, da dieser im Rahmen der Behandlung der fehlenden Werte explizit modelliert werden muss. Die entsprechenden Methoden wurden aus theoretischer und praktischer Sicht bislang wenig erforscht und ein wesentliches Ziel der vorliegenden Arbeit ist es, diese vorhandene Lücke zu schließen. Anhand eines methodischen Vergleichs der Verfahren sollen dabei Erkenntnisse gewonnen werden, die eine Bewertung im Hinblick auf ihre Anwendungsmöglichkeiten und die Verlässlichkeit ihrer Ergebnisse erlauben. Ferner wird in der Arbeit ein besonderer Wert auf die konkrete Umsetzung der Methoden gelegt, da sich diese als kompliziert erweist und für das jeweilige Anwendungsproblem spezifisch durchzuführen ist.

Aufgrund der Unkenntnis der fehlenden Werte beruhen die Verfahren zur Behandlung von nicht ignorierbaren Ausfallmechanismen ebenfalls auf Annahmen, die anhand der beobachteten Daten nicht überprüft werden können. In der vorliegenden Arbeit werden sowohl für diskrete als auch stetige Variablen Ansätze präsentiert, die eine Datenanalyse unter verschiedenen Annahmen über den Mechanismus ermöglichen und somit diese Problematik abschwächen. Durch diese Vorgehensweise wird gleichzeitig eine Beurteilung der Sensitivität von Schätzungen vorgenommen, womit der Einfluss der Annahmen auf die statistische Auswertung für den jeweiligen Anwendungsfall sichtbar wird.

---

[4] Vgl. Lillard et al. (1986); Rässler (2000), S. 67; Esser et al. (1989); Groves (1989), S. 201ff. Die Problematik wurde ebenfalls in Studien zum Wahlverhalten festgestellt, vgl. Smith et al. (1999); S. 563.

[5] Vgl. De Heer (1999), S. 129ff.; Esser et al. (1989), S. 99f.; Schnell (1997), S. 84f.

Ein weiterer Schwerpunkt der Arbeit befasst sich mit der Frage, inwieweit die in der Praxis verbreiteten Verfahren, welche auf einem durch die beobachteten Werte erklärbaren Datenausfall beruhen, die Güte von Schätzungen negativ beeinflussen, falls der Ausfallmechanismus nicht ignorierbar ist. In diesem Kontext werden Simulationsstudien durchgeführt, die unter verschiedenen Datenkonstellationen Aufschluss über das Ausmaß der Verzerrung von Schätzungen geben sollen.

Abschließend ist die Robustheit von Ansätzen, die auf einem bestimmten nicht ignorierbaren Ausfallmechanismus basieren, zu überprüfen, indem von deren zugrunde liegenden Unabhängigkeitsannahmen in mehreren Datensimulationen abgewichen wird. Auf der einen Seite sollen diese Simulationen die aus methodischer Sicht abgeleiteten Erkenntnisse unterstützen und andererseits zusätzliche Informationen über den Anwendungsbereich der Verfahren liefern.

1.2 Aufbau der Arbeit

Im Kontext der Behandlung von fehlenden Werten ist zunächst eine Abgrenzung der für die Problemstellung relevanten Variablen erforderlich. Neben dieser Festlegung werden die grundlegenden Annahmen, auf denen die nachfolgenden Ausführungen aufbauen, in **Kapitel 2** präzisiert. Ausgehend von dieser Vorbetrachtung werden in **Kapitel 3** die Ausfallmechanismen anhand ihrer Abhängigkeit von den beobachteten und fehlenden Daten klassifiziert. In dem Zusammenhang erfolgt eine kritische Betrachtung der beiden Annahmen, dass der Datenausfall zufällig respektive durch die beobachteten Werte erklärbar ist. Die Einteilung der Mechanismen bildet die Grundlage für die Definition der Ignorierbarkeit in **Kapitel 4,** die sowohl für die frequentistische als auch die bayesianische Theorie formalisiert wird. Beiden Sichtweisen folgend wird nachgewiesen, dass eine Erfüllung dieses Kriteriums die Möglichkeit eröffnet, gültige statistische Schlussfolgerungen aus den beobachteten Daten ziehen zu können.

Die Ignorierbarkeit des Ausfallmechanismus ist die zentrale Voraussetzung für die in **Kapitel 5** diskutierten likelihood-basierten Verfahren zur Behandlung von fehlenden Werten, deren Methodik und Umsetzung im Vordergrund der Ausführungen stehen. Weiterhin erfolgt in diesem Kapitel eine kritische Betrachtung von Verfahren, die auf der Annahme beruhen, dass der Datenausfall zufällig ist. Dabei werden insbe-

sondere die methodischen Nachteile dieser Methoden offen gelegt, die aufgrund ihrer einfachen Realisierbarkeit in der Praxis verbreitet sind.

Der Thematik der Arbeit entsprechend werden in **Kapitel 6** Methoden erörtert, die eine Behandlung von fehlenden Werten auch in Fällen ermöglichen, in denen das Kriterium der Ignorierbarkeit des Ausfallmechanismus nicht erfüllt ist. Einführend werden zwei Verfahren diskutiert, die bei der Betrachtung von einer diskreten bzw. zwei binären Variablen zur Anwendung kommen können. Diese Beschränkungen werden im Weiteren durch die Darstellung von allgemeinen Ansätzen (Selection Modelle und Pattern-Mixture Modelle) aufgehoben. Neben der konkreten Umsetzung dieser Modelle ist der methodische Vergleich der beiden Ansätze ein zentraler Bestandteil der Ausführungen in diesem Abschnitt.

Die im **Kapitel 7** beschriebenen Simulationsstudien befassen sich vornehmlich mit der Frage, inwieweit die in den vorangegangenen Kapiteln diskutierten Verfahren zu verzerrten Parameterschätzungen führen können, falls deren zugrunde liegenden Annahmen nicht erfüllt sind. Die Ergebnisse der Betrachtungen, die sowohl für den diskreten als auch den stetigen Fall erfolgen, münden in einer wertenden Gegenüberstellung der Behandlungsverfahren. In **Kapitel 8** werden die Erkenntnisse der Arbeit zusammengefasst, und es wird ein kurzer Ausblick auf die weiteren Forschungsmöglichkeiten gegeben.

2 Grundlagen

In den letzten Jahrzehnten findet die Problematik von fehlenden Werten eine verstärkte Beachtung, die sich in einer unüberschaubaren Zahl an Publikationen zu diesem Thema zeigt. Dabei ist im Rahmen von empirischen Untersuchungen unumstritten, dass sich bereits im Vorfeld der Erhebung mit dieser Problematik beschäftigt werden muss. So kann beispielsweise in Befragungen durch die eindeutige Formulierung der Fragen eine mögliche Ausfallursache weitgehend vermieden werden.[6] Aus der statistischen Literatur geht hervor, dass durch eine erneute Befragung von zufällig ausgewählten, nicht antwortenden Personen die Problematik des Datenausfalls deutlich verringert werden kann.[7] Eine Umsetzung dieser Strategie ist jedoch aus den unterschiedlichsten Gründen (z.B. Aufwand der Nachbefragung, Anonymitätsgründe) häufig nicht möglich, so dass die Notwendigkeit besteht, eine statistische Auswertung von lückenhaftem Datenmaterial ohne zusätzliche Informationen über die unbeobachteten Werte durchzuführen. Die Ergebnisse *einer* Erhebung, bei der die interessierenden Merkmale der einzelnen Untersuchungseinheiten nicht vollständig beobachtet werden konnten, sind der Ausgangspunkt der Arbeit.

Vor diesem Hintergrund ist eine entsprechende Notation einzuführen, die eine statistische Behandlung der Problematik allgemein erlaubt. Die Grundlage der folgenden Ausführungen bildet eine Untersuchung von k verschiedenen Merkmalen, zu deren Zweck eine einfache Stichprobe vom Umfang n aus der Grundgesamtheit gezogen wird. Die einzelnen Merkmale sind als Zufallsvariablen Y_j ($j = 1,...,k$) zu interpretieren, die zu einer k-dimensionalen Zufallsvariable $\underline{Y} = (Y_1,...,Y_k)$ mit den Realisationen $\underline{y} = (y_1,...,y_k)$ zusammengefasst werden können. Dementsprechend seien $(\underline{Y}_1,...,\underline{Y}_n)$ n mehrdimensionale Zufallsvariablen, die annahmegemäß unabhängig und identisch verteilt sind und die Realisationen $(\underline{y}_1,...,\underline{y}_n)$ mit $\underline{y}_i = (y_{i1},..., y_{ik})$ ($i = 1,..., n$) aufweisen. Die Werte $(\underline{y}_1,...,\underline{y}_n)$ definieren die Datenmatrix $\mathbf{y}$, welche eine Realisation der Zufallsmatrix $\mathbf{Y}$ ist und mit einer Wahrscheinlichkeit $P(\mathbf{y})$ beobachtet wird.

---

[6] Vgl. Schnell (1986), S. 24ff.; Bankhofer/Praxmarer (1998), S. 109; Runte (1999), S. 2.
[7] Vgl. Graham/Donaldson (1993), S. 124ff.; Glynn et al. (1993), S. 984; Huisman et al. (1999), S. 47ff.

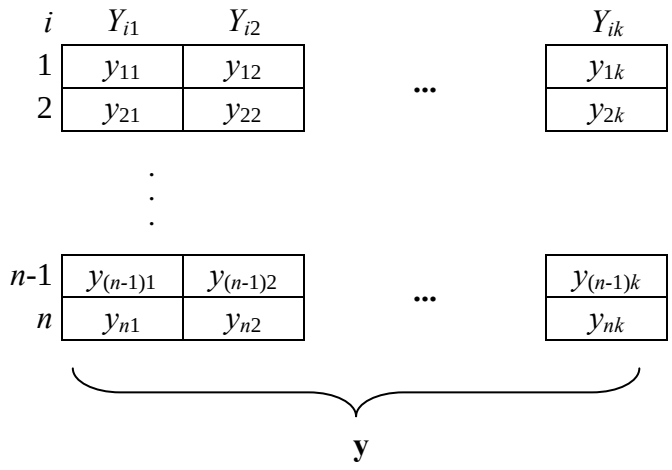


Abbildung 2.1: Darstellung der Datenmatrix **y**

Im Folgenden wird angenommen, dass die mehrdimensionale Zufallsvariable $\underline{Y}$ einen bestimmten Verteilungstyp besitzt und die Verteilung von $\underline{Y}$ durch einen Parametervektor θ vollständig beschrieben wird. Dementsprechend wird für die Verteilung von $\underline{Y}$ die Schreibweise $P_\theta(\underline{Y})$ eingeführt, welche diesen für die weiteren Ausführungen wesentlichen Aspekt unterstreichen soll. Folglich kann die Wahrscheinlichkeit bzw. Dichte von **y** ebenfalls durch θ beschrieben werden und entsprechend wird diese im Weiteren mit $P_\theta(\mathbf{y})$ bezeichnet.

Beispiel 2.1:

Die Zufallsvariable $\underline{Y}$ sei bivariat normalverteilt mit Erwartungswert $\mu = (\mu_1, \mu_2)$ und Kovarianzmatrix Σ. Für die Dichte der Datenmatrix **y** gilt aufgrund der Unabhängigkeit der Beobachtungen von $\underline{Y}$:

$$P_\theta(\mathbf{y}) = \prod_{i=1}^{n} P_\theta(\underline{y}_i) \qquad \theta = (\mu, \Sigma) \tag{2.1}$$

Ein wesentliches Ziel statistischer Analysen besteht darin, ausgehend von den beobachteten Daten Rückschlüsse über den unbekannten Parametervektor θ zu ziehen. Ist

die Datenmatrix **y** vollständig bekannt, so ist die Schätzung von θ durch gängige statistische Verfahren, wie z.B. die Maximum-Likelihood-Methode, möglich. Im Kontext der Problematik von fehlenden Werten sind einzelne Ausprägungen der Zufallsvariablen Y_j ($j = 1,...,k$) unbekannt, wodurch diese Verfahren nicht ohne weiteres angewendet werden können. Um den Datenausfall in **y** zu berücksichtigen, wird eine Indikatormatrix **R** eingeführt, welche die Elemente

- $R_{ij} = 0$, falls die Ausprägung des Merkmals j in Datensatz i - y_{ij} - beobachtet wurde, bzw.
- $R_{ij} = 1$, falls die Ausprägung des Merkmals j in Datensatz i nicht beobachtet wurde,

enthält. Die Indikatormatrix **R** ist als $n \cdot k$-elementiger Zufallsvektor zu interpretieren, der aus den Zufallsvariablen $(\underline{R}_1,...,\underline{R}_n)$ mit $\underline{R}_i = (R_{i1},...,R_{ik})$ $(i = 1,...,n)$ besteht. Die Zufallsvariablen $(\underline{R}_1,...,\underline{R}_n)$ sind annahmegemäß unabhängig und identisch verteilt und weisen die Realisationen $\underline{r}_i = (r_{i1},...,r_{ik})$ $(i = 1,...,n)$ in der Stichprobe auf. Ausgehend von der iid-Eigenschaft der Stichprobe wird weiterhin die Zufallsvariable $\underline{R} = (R_1,...,R_k)$ mit den Realisationen $\underline{r} = (r_1,...,r_k)$ definiert, welche die gleiche Verteilung wie die Zufallsvariablen $(\underline{R}_1,...,\underline{R}_n)$ besitzt.

i	Y_{i1}	Y_{i2}
1	5	7
2	3	?
3	?	2
4	?	?

i	R_{i1}	R_{i2}
1	0	0
2	0	1
3	1	0
4	1	1

r

Abbildung 2.2: Beispiel für die Realisation **r** der Indikatormatrix **R** bei einer zweidimensionalen Zufallsvariable $\underline{Y}$ (? = fehlender Wert)

Ein allgemeines Ziel der statistischen Verfahren zur Behandlung von fehlenden Werten ist es, ausgehend von einer Faktorisierung der gemeinsamen Wahrscheinlichkeit bzw. Dichte von **y** und **r** den interessierenden Parameter θ zu schätzen. Hierzu wird im Folgenden vorausgesetzt, dass die (Nicht-)Beobachtung von Werten ausschließ-

lich auf die Ausprägungen in der Datenmatrix **y** zurückzuführen ist. Dieser Zusammenhang wird durch die bedingte Verteilung $P(\mathbf{R} \mid \mathbf{y})$ formalisiert, welche allgemein als Ausfall- bzw. Antwortmechanismus bezeichnet wird.[8] Darüber hinaus besteht die Annahme, dass der Ausfallmechanismus durch einen Parameter ψ beschrieben wird und nicht vom Parameter θ abhängig ist:

$$P_{\psi,\theta}(\mathbf{R} \mid \mathbf{y}) = P_{\psi}(\mathbf{R} \mid \mathbf{y}) \tag{2.2}$$

Beispiel 2.2:

Die Ausfallwahrscheinlichkeit einer univariat normalverteilten Zufallsvariable $\underline{Y}$

$$\underline{Y} \sim N(\mu, \sigma^2), \qquad \theta = (\mu, \sigma)$$

sei von deren Realisation $\underline{y}$ in der folgenden Weise abhängig:

$$P_{\psi}(\underline{R} = 1 \mid \underline{Y} = \underline{y}) = \Phi(\psi \cdot \underline{y})$$

($\Phi(\cdot)$: Verteilungsfunktion der Standardnormalverteilung)

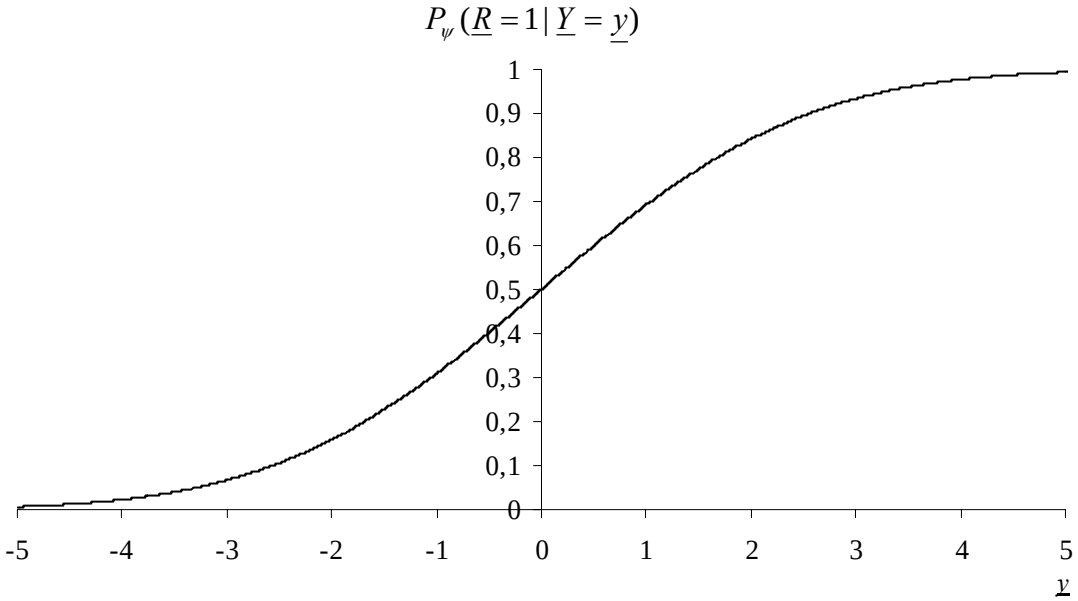


Abbildung 2.3: Bedingte Ausfallwahrscheinlichkeit $P_{\psi}(\underline{R} = 1 \mid \underline{Y} = \underline{y}) = \Phi(\psi \cdot \underline{y})$ für $\psi = 0{,}5$

[8] Vgl. Rässler (2000), S. 70.

Da sowohl die Zufallsvariablen ($\underline{Y}_1, ..., \underline{Y}_n$) als auch ($\underline{R}_1, ..., \underline{R}_n$) unabhängig und identisch verteilt sind, gilt für den Ausfallmechanismus $P(\mathbf{R} \mid \mathbf{y})$:

$$\begin{aligned} P(\mathbf{R} \mid \mathbf{y}) &= \prod_{i=1}^{n} P_\psi(\underline{R}_i = \underline{r}_i \mid \underline{Y}_i = \underline{y}_i) \\ &= \prod_{i=1}^{n} (1-\underline{r}_i)\, P_\psi(\underline{R}_i = 0 \mid \underline{Y}_i = \underline{y}_i) + \underline{r}_i\ P_\psi(\underline{R}_i = 1 \mid \underline{Y}_i = \underline{y}_i) \\ &= \prod_{i=1}^{n} (1-\underline{r}_i)\left[1 - \Phi(\psi \cdot \underline{y}_i)\right] + \underline{r}_i\ \Phi(\psi \cdot \underline{y}_i) \end{aligned}$$

Der Ausfallmechanismus $P(\mathbf{R} \mid \mathbf{y})$ ist nicht vom Parametervektor θ der Verteilung von $\underline{Y}$ abhängig und wird allein durch den Parameter ψ beschrieben. Die in (2.2) getroffene Annahme ist in diesem Beispiel erfüllt.

Unter der Gültigkeit der in diesem Kapitel getroffenen Annahmen lässt sich die gemeinsame Wahrscheinlichkeit bzw. Dichte von **y** und **r** als Produkt der marginalen Wahrscheinlichkeit (Dichte) von **y** und der Ausfallwahrscheinlichkeit $P_\psi(\mathbf{r} \mid \mathbf{y})$ darstellen:

$$P_{\theta,\psi}(\mathbf{r}, \mathbf{y}) = P_\theta(\mathbf{y}) P_\psi(\mathbf{r} \mid \mathbf{y}) \tag{2.3}$$

Im Kontext des Datenausfalls ist es weiterhin notwendig, die beobachteten und fehlenden Elemente in der Matrix **y** getrennt zu definieren. Dementsprechend wird mit $\mathbf{y}_{obs} = (\underline{y}_{obs,1}, ..., \underline{y}_{obs,n})^T$ der beobachtete und mit $\mathbf{y}_{mis} = (\underline{y}_{mis,1}, ..., \underline{y}_{mis,n})^T$ der unbeobachtete Teil von $\mathbf{y} = (\mathbf{y}_{obs}, \mathbf{y}_{mis})$ bezeichnet. Die in $\mathbf{y}_{obs}$ und $\mathbf{y}_{mis}$ zusammengefassten, beobachteten und nicht beobachteten Elemente der Datenmatrix **y** sind als Realisationen von $\mathbf{Y}_{obs}$ und $\mathbf{Y}_{mis}$ zu interpretieren.[9]

[9] Es gilt somit $\mathbf{Y} = (\mathbf{Y}_{obs}, \mathbf{Y}_{mis})$.

i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}
1	5	7	0	?
2	3	4	?	?
3	2	2	?	?
4	1	?	?	?

Abbildung 2.4: Darstellung von $\mathbf{y}_{obs}$ (dunkelgrau) und $\mathbf{y}_{mis}$ (hellgrau) in einer Datenmatrix $\mathbf{y}$ (? = fehlender Wert)

Durch diese Unterteilung der Datenmatrix $\mathbf{y}$ sind die Wahrscheinlichkeiten bzw. Dichten von $\mathbf{y}_{obs}$ und $\mathbf{y}_{mis}$ nur von dem Parametervektor θ abhängig und werden mit $P_\theta(\mathbf{y}_{obs})$ beziehungsweise $P_\theta(\mathbf{y}_{mis})$ bezeichnet. Ist die mehrdimensionale Zufallsvariable $\underline{Y}$ stetig, so ist die Dichte von $\mathbf{y}_{obs}$ formal durch Integration der Dichtefunktion $P_\theta(\mathbf{y})$ über die fehlenden Werte zu erhalten:[10]

$$P_\theta(\mathbf{y}_{obs}) = \int P_\theta(\mathbf{y})\, d\mathbf{y}_{mis} = \prod_{i=1}^{n} \int P_\theta(\underline{y}_{obs,i}, \underline{y}_{mis,i})\, d\underline{y}_{mis,i} = \prod_{i=1}^{n} P_\theta(\underline{y}_{obs,i}) \tag{2.4}$$

Im diskreten Fall ist die Wahrscheinlichkeitsfunktion $P_\theta(\mathbf{y})$ über alle Werte der unbeobachteten Variablen zu summieren:

$$P_\theta(\mathbf{y}_{obs}) = \prod_{i=1}^{n} \sum_{\underline{y}_{mis,i}} P_\theta(\underline{y}_{obs,i}, \underline{y}_{mis,i}) = \prod_{i=1}^{n} P_\theta(\underline{y}_{obs,i}) \tag{2.5}$$

Zur Korrektur des Datenausfalls sind zahlreiche Methoden bekannt, die unter verschiedenen Bedingungen des Ausfalls zur Anwendung kommen können. Die Ausführungen zur Methodik dieser Verfahren erfordern zunächst eine Darstellung und Einteilung ihrer Anwendungsvoraussetzungen.

[10] Es gelte wiederum die iid-Eigenschaft der Stichprobe.

3 Klassifikation der Ausfallmechanismen nach Rubin

Die Behandlung von fehlenden Werten setzt die Kenntnis über den Zusammenhang von der (Nicht-)Beobachtung der Merkmale und den Merkmalsausprägungen in einer Stichprobe voraus. Die genaue Spezifizierung dieses als Ausfallmechanismus bezeichneten Zusammenhangs ist im Allgemeinen nicht möglich; es können jedoch Unabhängigkeitsannahmen getroffen werden, unter deren Gültigkeit die Behandlung mit entsprechenden Methoden zu unverzerrten Schätzern bei der anschließenden Datenanalyse führt. Die implizit durch die Annahmen erfolgte, grundsätzliche Einteilung der Ausfallmechanismen beruht auf der Unterscheidung, ob das Fehlen von Variablen zufällig oder von Merkmalen derselben Untersuchungseinheit, insbesondere auch von dem fehlenden Merkmal selbst, abhängig ist. Die Ausfallmechanismen lassen sich somit nach Rubin[11] in drei Gruppen klassifizieren:

(I) *„Missing Completely at Random" (MCAR)*

Die den Ausfallmechanismus beschreibende bedingte Verteilung von **R** gegeben **y** entspricht der Randverteilung von **R**:

$$P_{\psi}(\mathbf{R} \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) = P_{\psi}(\mathbf{R}) \tag{3.1}$$

Die Wahrscheinlichkeit für das Fehlen eines Wertes ist unabhängig von den (beobachteten und unbeobachteten) Ausprägungen aller Merkmale bei einer Untersuchungseinheit. Weiterhin hängt der Datenausfall auch nicht von den Merkmalswerten anderer Einheiten ab, so dass in diesem Zusammenhang auch von einem zufälligen bzw. unsystematischen Fehlen der Werte gesprochen wird.[12] Die MCAR-Annahme ist nur in seltenen Fällen gerechtfertigt und wird von einigen Autoren sogar als „unrealistisch" bei empirischen Datenbeständen bezeichnet.[13]

11 Vgl. Rubin (1976).
12 Vgl. Rässler (2000), S. 65.
13 Vgl. Schafer/Olsen (1998), S. 11.

Beispiel 3.1:

Die eindimensionale Zufallsvariable $\underline{R}_i$ $(1 \le i \le n)$, die den Datenausfall einer ebenfalls univariaten Zufallsvariable $\underline{Y}_i$ in Untersuchungseinheit *i* beschreibt, sei Bernoulli-verteilt mit Parameter ψ $\left(\underline{R}_i \sim B(1,\psi)\right)$. Allgemein gilt bei einer iid-Stichprobe vom Umfang *n* für den Ausfallmechanismus $P_\psi(\mathbf{R} \mid \mathbf{y})$:

$$P_\psi(\mathbf{R} \mid \mathbf{y}) = \prod_{i=1}^{n} P_\psi(\underline{R}_i = \underline{r}_i \mid \underline{Y}_i = \underline{y}_i) \tag{3.2}$$

Die Annahme einer Bernoulli-verteilten Zufallsvariable $\underline{R}_i$ impliziert deren Unabhängigkeit von der Zufallsvariable $\underline{Y}_i$:

$$P_\psi(\underline{R}_i = 1 \mid \underline{Y}_i = \underline{y}_i) = P_\psi(\underline{R}_i = 1) = \psi$$

$$P_\psi(\underline{R}_i = 0 \mid \underline{Y}_i = \underline{y}_i) = P_\psi(\underline{R}_i = 0) = 1 - \psi$$

Durch Einsetzen dieser Unabhängigkeitsbeziehung in (3.2) kann gezeigt werden, dass die Bedingung in (3.1) erfüllt und somit der Ausfallmechanismus vom Typ MCAR ist:

$$P_\psi(\mathbf{R} \mid \mathbf{y}) = \prod_{i=1}^{n} P_\psi(\underline{R}_i = \underline{r}_i \mid \underline{Y}_i = \underline{y}_i) = \prod_{i=1}^{n} P_\psi(\underline{R}_i = \underline{r}_i) = P_\psi(\mathbf{R}) \tag{3.3}$$

(II) *„Missing at Random" (MAR)*

Das Fehlen von Werten ist bei diesem Ausfallmechanismus lediglich von den beobachteten und *nicht* von den unbeobachteten Daten abhängig:

$$P_\psi(\mathbf{R} \mid \mathbf{y}) = P_\psi(\mathbf{R} \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) = P_\psi(\mathbf{R} \mid \mathbf{y}_{obs}) \tag{3.4}$$

Die MAR-Annahme ist weniger restriktiv als die MCAR-Annahme, da die Werte bei diesem Mechanismus nicht (vollständig) zufällig fehlen. Allerdings kann der Datenausfall allein aus den beobachteten Daten erklärt werden, so dass die fehlenden Werte als Zufallsstichprobe gegeben die beobachteten Realisationen zu betrachten sind.

In diesem Zusammenhang wird in der Literatur auch von bedingt zufälligem Datenausfall gesprochen.[14]

Beispiel 3.2:

Von zwei binären Zufallsvariablen Y_1 und Y_2 sei lediglich Y_2 vom Datenausfall betroffen. Weiterhin sei bekannt, dass die Ausfallwahrscheinlichkeit bezüglich Y_2 ausschließlich von den Realisationen von Y_1 abhängig ist:

$$P_\psi(R_{i2} = 1 \mid Y_{i1} = 0) = \psi_1$$

$$P_\psi(R_{i2} = 1 \mid Y_{i1} = 1) = \psi_2 \qquad (i = 1,...,n)$$

Aufgrund der iid-Eigenschaft der Stichprobe vom Umfang n ist der Ausfallmechanismus $P_\psi(\mathbf{R} \mid \mathbf{y})$ mit $\psi = (\psi_1, \psi_2)$ nur von den beobachteten Daten $\mathbf{y}_{obs}$ abhängig:

$$P_\psi(\mathbf{R} \mid \mathbf{y}) = \prod_{i=1}^{n} P_\psi(\underline{R}_i = \underline{r}_i \mid \underline{Y}_i = \underline{y}_i) = \prod_{i=1}^{n} P_\psi(R_{i1} = 0, R_{i2} = r_{i2} \mid Y_{i1} = y_{i1})$$

$$= \prod_{i=1}^{n} P_\psi(R_{i1} = 0, R_{i2} = r_{i2} \mid \underline{y}_{obs,i}) = P_\psi(\mathbf{R} \mid \mathbf{y}_{obs})$$

Nach der Definition in (3.4) liegt somit ein Ausfallmechanismus vom Typ MAR vor.

Die Gültigkeit von MAR kann in der Praxis nicht nachgewiesen werden, da die Überprüfung der notwendigen Bedingung (3.4) aufgrund der unbekannten Daten $\mathbf{y}_{mis}$ nicht möglich ist. Hierzu besteht bei mehreren Autoren[15] die Auffassung, dass die Problematik der Plausibilität von MAR insbesondere bei umfangreichen, multivariaten Datenbeständen zu vernachlässigen ist, da in diesen Fällen häufig der Zusammenhang zwischen Ausfallsmechanismus und den Werten der fehlenden Variablen hinreichend durch die beobachteten Daten erklärt werden kann. Darüber hinaus wird argumentiert, dass die Behandlung von fehlenden Werten mittels Methoden, die einen Ausfallmechanismus vom Typ MAR voraussetzen, generell bei Untersuchungen

[14] Vgl. Rässler (2000), S. 65.
[15] Vgl. Schafer (1997), S. 27; David et al. (1986); Schnell (2002) S. 12.

von empirischen Datenbeständen gerechtfertigt ist, da möglicherweise vorhandene Verletzungen der MAR-Annahme keine wesentlichen Auswirkungen auf die Analyseergebnisse zur Folge haben.[16] Unter anderem aus diesen Gründen ist es zu erklären, dass sich die statistische Literatur vornehmlich diesem Ausfalltyp widmet und hierfür zahlreiche Behandlungsmethoden existieren.

(III) *„Missing Not at Random" (MNAR)*

Bei diesem Mechanismus hängt die Ausfallwahrscheinlichkeit – im Gegensatz zu „Missing at Random" und „Missing Completely at Random" – auch von den fehlenden Werten ab:

$$P_\psi(\mathbf{R} \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \neq P_\psi(\mathbf{R} \mid \mathbf{y}_{obs}) \tag{3.5}$$

Beispiel 3.3:

Ausgehend vom Beispiel 3.2 sei im Weiteren die Ausfallwahrscheinlichkeit nur von den Realisationen der Zufallsvariable Y_2 abhängig:

$$P_\psi(R_{i2} = 1 \mid Y_{i2} = 0) = \psi_1$$

$$P_\psi(R_{i2} = 1 \mid Y_{i2} = 1) = \psi_2 \qquad (i = 1,...,n)$$

Demnach ist der Ausfallmechanismus nicht allein aus den beobachteten Daten erklärbar und somit vom Typ MNAR:

$$P_\psi(\mathbf{R} \mid \mathbf{y}) = \prod_{i=1}^{n} P_\psi(\underline{R}_i = \underline{r}_i \mid \underline{Y}_i = \underline{y}_i) = \prod_{i=1}^{n} P_\psi(R_{i1} = 0, R_{i2} = r_{i2} \mid Y_{i2} = y_{i2})$$

$$= \prod_{i=1}^{n} P_\psi(R_{i1} = 0, R_{i2} = r_{i2} \mid \underline{y}_{obs,i}, \underline{y}_{mis,i}) \neq P_\psi(\mathbf{R} \mid \mathbf{y}_{obs})$$

Die Behandlung von fehlenden Werten unter MNAR erfordert im Allgemeinen die Kenntnis der bedingten Verteilung $P_\psi(\mathbf{R} \mid \mathbf{y}_{obs}, \mathbf{y}_{mis})$, die aufgrund der Abhängigkeit von den unbeobachteten Werten $\mathbf{y}_{mis}$ jedoch in den meisten Fällen unbekannt ist. Da

[16] Vgl. Schafer/Graham (2002), S. 152.

eine genaue Spezifizierung der bedingten Verteilung von **R** häufig nicht möglich ist, werden für den Datenausfall vom Typ MNAR geeignete Behandlungsmethoden in der Praxis nur angewandt, falls inhaltliche Gründe der MAR-Annahme grundlegend widersprechen. In diesen letztgenannten Fällen werden oft mehrere unterschiedliche, plausible Verteilungen für den Datenausfall unterstellt, um anschließend eine Sensitivitätsbetrachtung bei der Datenanalyse durchzuführen und somit der Unsicherheit gegenüber dem genauen Ausfallmechanismus Rechnung zu tragen.[17]

[17] Vgl. Allison (2002), S. 78; Toutenburg et al. (2004), S. 34.

4 Ignorierbarkeit des Ausfallmechanismus

Um die Verteilung von $\underline{Y}$ bzw. den interessierenden Parametervektor θ aus den unvollständigen Datenbeständen bestimmen zu können, wurde von Rubin (1976) die „Ignorability"-Eigenschaft von Ausfallmechanismen definiert. Bei Vorliegen dieser Eigenschaft ist es nicht notwendig, die bedingte Verteilung von **R** gegeben **y** zu spezifizieren, um entsprechende Methoden zum Schätzen von θ anwenden zu können. Der Ausfallmechanismus kann somit für diese Schätzung ignoriert werden. Da Inferenzen bezüglich θ sowohl aus frequentistischer als auch bayesianischer Sicht erfolgen können, ist die Ignorierbarkeit für die beiden Theorien unterschiedlich zu definieren. Die folgenden Ausführungen beziehen sich zunächst auf den frequentistischen Ansatz.

Definition 4.1:

Innerhalb der frequentistischen Theorie wird ein Ausfallmechanismus als ignorierbar bezeichnet, falls gilt:[18]

(I) Der Ausfallmechanismus ist vom Typ MCAR oder MAR.

(II) Der gemeinsame, mit $\mathbf{\Omega}_{\theta,\psi}$ bezeichnete Parameterraum von (θ, ψ) ist das karthesische Kreuzprodukt des mit $\mathbf{\Omega}_{\theta}$ benannten Parameterraums von θ und des durch $\mathbf{\Omega}_{\psi}$ symbolisierten Parameterraums von ψ:

$$\mathbf{\Omega}_{\theta,\psi} = \mathbf{\Omega}_{\theta} \times \mathbf{\Omega}_{\psi}$$

Aus der Ignorierbarkeit des Ausfallmechanismus können Folgerungen für die Wahrscheinlichkeits- bzw. Dichtefunktion von den beobachteten Werten $\mathbf{y}_{obs}$ und **r** abgeleitet werden. Bei einer stetigen Zufallsvariable Y ist die Dichtefunktion der beobachteten Daten durch Integration von $P_{\theta,\psi}(\mathbf{r}, \mathbf{y})$ über die fehlenden Werte $\mathbf{y}_{mis}$ zu erhalten:[19]

---

[18] Vgl. Rubin (1976), S. 586.
[19] Im Fall von diskreten Daten ist das Integralzeichen durch ein Summenzeichen zu ersetzen.

$$P_{\theta,\psi}(\mathbf{r},\mathbf{y}_{obs}) = \int P_{\theta,\psi}(\mathbf{r},\mathbf{y})\; d\mathbf{y}_{mis} \tag{4.1}$$

Die Faktorisierung der gemeinsamen Dichte von **r** und **y** führt unter den in Kapitel 2 genannten Voraussetzungen zu

$$P_{\theta,\psi}(\mathbf{r},\mathbf{y}_{obs}) = \int P_{\psi}(\mathbf{r} \mid \mathbf{y})\; P_{\theta}(\mathbf{y})\; d\mathbf{y}_{mis}\;. \tag{4.2}$$

Bei Erfüllung der MAR-Annahme $P_{\psi}(\mathbf{R} \mid \mathbf{y}) = P_{\psi}(\mathbf{R} \mid \mathbf{y}_{obs})$ gilt daher:[20]

$$\begin{aligned} P_{\psi,\theta}(\mathbf{r},\mathbf{y}_{obs}) &= \int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})\; P_{\theta}(\mathbf{y})\; d\mathbf{y}_{mis} \\ &= P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs}) \int P_{\theta}(\mathbf{y}_{obs},\mathbf{y}_{mis})\; d\mathbf{y}_{mis} \\ &= P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})\; P_{\theta}(\mathbf{y}_{obs}) \end{aligned} \tag{4.3}$$

Ist die „Ignorability"-Voraussetzung (II) in Definition 4.1 erfüllt, können alle Folgerungen für θ aus der Likelihood-Funktion

$$L(\mathbf{y}_{obs} \mid \theta) \propto P_{\theta}(\mathbf{y}_{obs})$$

gezogen werden, da der Parameterwert von θ mit allen möglichen Parameterwerten von ψ kombinierbar und keine Information über θ in $P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})$ enthalten ist. Dies impliziert, dass bei ignorierbarem Ausfallmechanismus die Maximum-Likelihood-Schätzung für den Parametervektor θ der Maximierung von $L(\mathbf{y}_{obs} \mid \theta)$ entspricht. Little/Rubin (2002) weisen in diesem Zusammenhang darauf hin, dass innerhalb der frequentistischen Sichtweise auch bei einer Verletzung der Bedingung (II) in Definition 4.1 korrekte Inferenzen bezüglich des Parametervektors θ aus der Likelihood-Funktion $L(\mathbf{y}_{obs} \mid \theta)$ gezogen werden können, allerdings führt die Ignorierung des Ausfallmechanismus in (4.3) dann zu einem Verlust an Effizienz bei der Schätzung von θ.[21] Weiterhin wird von Allison (2002) die Auffassung vertreten, dass ein Zusammenhang der beiden Parametervektoren θ und ψ in praktischen Anwendungen unrealistisch ist.[22] Insofern ist die Voraussetzung (I), die das Vorliegen eines Aus-

[20] Vgl. Schafer (1997), S. 12.
[21] Vgl. Little/Rubin (2002), S. 120; Wilhelm (1996), S. 76.
[22] Vgl. Allison (2002), S. 5.

fallmechanismus vom Typ MCAR bzw. MAR fordert, als der zentrale Bestandteil für die Definition der Ignorierbarkeit zu betrachten.

Ist die letztgenannte Voraussetzung nicht erfüllt, so sind keine Schlüsse für θ aus der Likelihood-Funktion $L(\mathbf{y}_{obs} \mid \theta)$ möglich. In diesen Fällen kann der Ausfallmechanismus bei der Maximum-Likelihood-Schätzung nicht vernachlässigt werden („non-ignorable“) und der Parametervektor θ ist aus der Likelihood-Funktion

$$L(\mathbf{y}_{obs}, \mathbf{r} \mid \theta, \psi) \propto P_{\theta,\psi}(\mathbf{y}_{obs}, \mathbf{r})$$

zu schätzen. Bei Datenausfall vom Typ MNAR ist der Ausfallmechanismus stets nicht ignorierbar.

Im Gegensatz zum frequentistischen Ansatz wird der bayesianischen Theorie zufolge der Parametervektor θ als Zufallsvariable aufgefasst, welche eine a priori Verteilung $P(\theta)$ besitzt. Die Inferenzen bezüglich θ werden bei vollständiger Beobachtung der Daten $\mathbf{y}$ aus der a posteriori Verteilung $P(\theta \mid \mathbf{y})$ gezogen, indem der Satz von Bayes angewendet wird:[23]

$$P(\theta \mid \mathbf{y}) = \frac{P(\theta)P(\mathbf{y} \mid \theta)}{P(\mathbf{y})} \tag{4.4}$$

Dabei entspricht $P(\mathbf{y} \mid \theta)$ der in Kapitel 2 definierten, durch den Parametervektor θ beschriebenen Wahrscheinlichkeit bzw. Dichte von $\mathbf{y}$ $\left(P(\mathbf{y} \mid \theta) = P_{\theta}(\mathbf{y})\right)$, während $P(\mathbf{y})$ durch die folgende Integration zu erhalten ist:

$$P(\mathbf{y}) = \int P(\mathbf{y} \mid \theta)P(\theta)d\theta \tag{4.5}$$

Innerhalb der bayesianischen Theorie wird ein Ausfallmechanismus als ignorierbar bezeichnet, falls gilt: [24]

[23] Vgl. Bamberg (1972), S. 101.
[24] Vgl. Little/Rubin (2002), S. 120.

Definition 4.2:

(I) Der Ausfallmechanismus ist vom Typ MCAR oder MAR.

(II) Die Parametervektoren θ und ψ sind a priori unabhängig, so dass die gemeinsame a priori Verteilung $P(\theta,\psi)$ das Produkt der beiden a priori Verteilungen von θ und ψ ist:

$$P(\theta,\psi) = P(\theta)P(\psi)$$

Die Definitionen der Ignorierbarkeit des Ausfallmechanismus in Definition 4.1 und Definition 4.2 unterscheiden sich in der jeweiligen Bedingung (II). Dabei wird nach der bayesianischen Theorie eine stärkere Annahme getroffen, da die Verschiedenheit der Parameterräume von θ und ψ, die nach der frequentistischen Definition der Ignorierbarkeit gefordert wird, eine notwendige Voraussetzung für die Unabhängigkeit der a priori Verteilungen von diesen beiden Parametervektoren ist.[25]

Im allgemeinen Fall der Unvollständigkeit von **y** sind bayesianische Inferenzen bezüglich θ aus der marginalen a posteriori Verteilung $P(\theta \mid \mathbf{y}_{obs}, \mathbf{r})$ durch Integration der a posteriori Verteilung $P(\theta,\psi \mid \mathbf{y}_{obs}, \mathbf{r})$ über den Parametervektor ψ zu ziehen:

$$P(\theta \mid \mathbf{y}_{obs}, \mathbf{r}) = \int P(\theta,\psi \mid \mathbf{y}_{obs}, \mathbf{r})\, d\psi \tag{4.6}$$

Im stetigen Fall gilt für die a posteriori Verteilung $P(\theta,\psi \mid \mathbf{y}_{obs}, \mathbf{r})$ nach dem Satz von Bayes:[26]

$$P(\theta,\psi \mid \mathbf{y}_{obs}, \mathbf{r}) = \frac{P_{\theta,\psi}(\mathbf{r}, \mathbf{y}_{obs})P(\theta,\psi)}{\int\int P_{\theta,\psi}(\mathbf{r}, \mathbf{y}_{obs})P(\theta,\psi)d\theta\, d\psi} \tag{4.7}$$

Da das Doppelintegral $\int\int P_{\theta,\psi}(\mathbf{r}, \mathbf{y}_{obs})P(\theta,\psi)d\theta\, d\psi$ als Normierungsfaktor betrachtet werden kann, ist die a posteriori Verteilung $P(\theta,\psi \mid \mathbf{y}_{obs}, \mathbf{r})$ proportional zum Produkt aus der Dichte $P_{\theta,\psi}(\mathbf{r}, \mathbf{y}_{obs})$ und der a priori Verteilung von θ und ψ:

---

[25] Vgl. Little/Rubin (2002), S. 120.
[26] Vgl. Schafer (1997), S. 17.

$$P(\theta,\psi \mid \mathbf{y}_{obs},\mathbf{r}) \propto P_{\theta,\psi}(\mathbf{r},\mathbf{y}_{obs})P(\theta,\psi) \tag{4.8}$$

Wie bereits in Formel (4.3) gezeigt gilt unter der MAR-Annahme für die Dichte der beobachteten Daten $P_{\theta,\psi}(\mathbf{r},\mathbf{y}_{obs}) = P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})\, P_{\theta}(\mathbf{y}_{obs})$, und durch Einsetzen in (4.8) erhält man

$$P(\theta,\psi \mid \mathbf{y}_{obs},\mathbf{r}) \propto P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P_{\theta}(\mathbf{y}_{obs})P(\theta,\psi). \tag{4.9}$$

Unter Voraussetzung (II) der bayesianischen Definition von Ignorierbarkeit, welche die Unabhängigkeit der a priori Verteilungen von θ und ψ postuliert, gilt für die marginale a posteriori Verteilung von θ:[27]

$$\begin{aligned} P(\theta \mid \mathbf{y}_{obs},\mathbf{r}) &= \int P(\theta,\psi \mid \mathbf{y}_{obs},\mathbf{r}) d\psi \\ &\propto \int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P_{\theta}(\mathbf{y}_{obs})P(\theta,\psi)\, d\psi \\ &\propto \int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P_{\theta}(\mathbf{y}_{obs})P(\theta)P(\psi)\, d\psi \\ &\propto P_{\theta}(\mathbf{y}_{obs})P(\theta)\int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P(\psi)\, d\psi \end{aligned} \tag{4.10}$$

Folglich gilt für die marginale a posteriori Verteilung von θ die Proportionalität

$$P(\theta \mid \mathbf{y}_{obs},\mathbf{r}) \propto P_{\theta}(\mathbf{y}_{obs})P(\theta), \tag{4.11}$$

da das Integral in (4.10) einen Faktor bildet, der nicht von θ abhängig ist. Aus (4.11) ist ersichtlich, dass die realisierte Indikatormatrix **r** nicht in die beiden Faktoren auf der rechten Seite der Formel einfließt, und somit ist auch die Proportionalität

$$P(\theta \mid \mathbf{y}_{obs}) \propto P_{\theta}(\mathbf{y}_{obs})P(\theta) \tag{4.12}$$

erfüllt.[28]

Ausgehend von der Definition der Likelihood-Funktion von θ gegeben die beobachteten Daten

$$L(\mathbf{y}_{obs} \mid \theta) \propto P_{\theta}(\mathbf{y}_{obs})$$

gilt entsprechend:

---

[27] Vgl. Schafer (1997), S. 17.
[28] Vgl. Schafer (1997), S. 17.

$$P(\theta \mid \mathbf{y}_{obs}) \propto L\ (\mathbf{y}_{obs} \mid \theta) P(\theta) \tag{4.13}$$

Die Formel (4.13) bildet die theoretische Grundlage für zahlreiche bayesianische Methoden zur Behandlung von fehlenden Werten.[29] Ist der Ausfallmechanismus ignorierbar im Sinne von Definition 4.2, so können bayesianische Inferenzen bezüglich θ aus der a posteriori Verteilung $P(\theta \mid \mathbf{y}_{obs})$ gezogen werden. Diese erfolgen beispielsweise durch die Maximierung des Produkts aus der Likelihood-Funktion $L(\mathbf{y}_{obs} \mid \theta)$ und der a priori Verteilung $P(\theta)$ bezüglich des Parametervektors θ. Bei großen Stichprobenumfängen ist die a priori Verteilung dabei von geringer Bedeutung für das Maximierungsproblem.[30]

Die „Ignorability"-Eigenschaft des Ausfallmechanismus bildet die theoretische Grundlage für die likelihood-basierten Behandlungsverfahren in Kapitel 5.3. Im Folgenden werden zunächst Verfahren behandelt, die aufgrund ihrer einfachen Umsetzbarkeit in der Praxis verbreitet sind. Insbesondere werden die methodischen Nachteile dieser Methoden herausgestellt, während anschließend Verfahren diskutiert werden, welche diese Problematik nicht aufweisen.

[29] Eine dieser Methoden, das Data-Augmentation Verfahren, wird in Kapitel 5.3.2 genauer erläutert.
[30] Vgl. Schafer/Olsen (1998), S. 10.

5 Behandlungsverfahren bei ignorierbaren Ausfallmechanismen

5.1 Traditionelle Verfahren

5.1.1 Vorbetrachtung

Zur Behandlung von fehlenden Werten werden in der statistischen Literatur zahlreiche Methoden aufgeführt, deren Anwendungen an bestimmte statistische Analyseziele (z.B. Schätzung von Regressionsparametern, Varianzschätzung) geknüpft sind.[31] Die in diesem Kapitel diskutierten Verfahren können demgegenüber unabhängig von der nachfolgenden Auswertung eingesetzt werden, jedoch sind diese entweder an die restriktive MCAR-Annahme (Eliminierungsverfahren, Mittelwertergänzung) oder an ein bestimmtes Ausfallmuster (Gewichtungsverfahren) gebunden.[32] Trotz der genannten Einschränkungen sind diese als traditionelle Methoden bezeichneten Verfahren in vielen statistischen Programmpaketen implementiert, so dass im Weiteren eine kritische Betrachtung dieser Behandlungsweisen erfolgt.

5.1.2 Eliminierungsverfahren

Eine häufig als Complete Case Analysis bezeichnete Methode zur Behandlung von fehlenden Werten besteht in der Vernachlässigung der unvollständigen Datensätze und einer anschließenden statistischen Analyse der vollständig beobachteten Untersuchungseinheiten. Die statistische Auswertung kann dabei durch die Anwendung von multivariaten Standardverfahren für vollständige Daten erfolgen. Durch die Complete Case Analysis wird insbesondere die Vergleichbarkeit von univariaten Statistiken gewährleistet, da die Analyse jeweils auf der gleichen Stichprobengröße basiert.[33]

Ein wesentlicher Nachteil dieser Verfahrensweise ist der zusätzliche Informationsverlust, der sich aufgrund der Nichtberücksichtigung von vorhandenen Beobachtungen in multivariaten Datensätzen, die im Extremfall nur einen fehlenden Wert aufweisen, ergibt. Weiterhin führt die Anwendung der Complete Case Analysis nur

---

[31] Vgl. Hübler (1986); Schnell (1986), S. 76ff.; Little/Rubin (2002), S. 28.
[32] Das Gewichtungsverfahren setzt zusätzlich die Erfüllung der MAR-Annahme voraus.
[33] Vgl. Little/Rubin (2002), S. 41; Runte (1999), S. 10.

dann zu unverzerrten Ergebnissen, wenn der Datenausfall zufällig im Sinne von MCAR ist und die vollständigen Datensätze eine repräsentative Stichprobe der Grundgesamtheit bilden.

Beispiel 5.1:

Es wird eine vollständige Datenmatrix **y**, bestehend aus den Realisationen von drei kardinal skalierten Zufallsvariablen Y_1, Y_2 und Y_3, betrachtet.

i	Y_{i1}	Y_{i2}	Y_{i3}
1	3	4	5
2	2	6	2
3	3	4	5
4	4	1	1
5	4	1	1
6	2	6	2

Die folgenden Werte der Zufallsvariablen seien beobachtet worden:

i	Y_{i1}	Y_{i2}	Y_{i3}
1	3	4	5
2	2	6	2
3	3	?	5
4	4	1	1
5	4	?	1
6	2	?	2

i	R_{i1}	R_{i2}	R_{i3}
1	0	0	0
2	0	0	0
3	0	1	0
4	0	0	0
5	0	1	0
6	0	1	0

Um die Gültigkeit der MCAR-Annahme überprüfen zu können, ist die bedingte Wahrscheinlichkeit $P_\psi(\mathbf{r} \mid \mathbf{y})$ sowie die Wahrscheinlichkeit $P_\psi(\mathbf{r})$ aus den relativen Häufigkeiten der Daten zu schätzen. Im Beispiel weist die Zufallsvariable Y_2 in drei Untersuchungseinheiten der Stichprobe vom Umfang $n = 6$ fehlende Werte auf, während aus der vollständigen Datenmatrix **y** und der Indikatormatrix **r** die Wahrscheinlichkeit $P(R_2 = 1 \mid \underline{Y} = \underline{y}) = 0{,}5$ für alle Ausprägungen von $\underline{Y}$ geschätzt wird.

Ausgehend von diesen Schätzungen entspricht die Wahrscheinlichkeit für die realisierte Indikatormatrix **r**

$$P_\psi(\mathbf{r}) = \prod_{i=1}^{6} P_\psi(\underline{r}_i) = \prod_{i=1}^{6} P_\psi(r_{i2}) = 0{,}5^6$$

der bedingten Wahrscheinlichkeit

$$P_\psi(\mathbf{r} \mid \mathbf{y}) = \prod_{i=1}^{6} P_\psi(\underline{r}_i \mid \underline{y}_i) = \prod_{i=1}^{6} P_\psi(r_{i2} \mid \underline{y}_i) = 0{,}5^6 \ ,$$

so dass von der Gültigkeit der MCAR-Annahme in diesem Beispiel ausgegangen werden kann. Die Anwendung der Complete Case Analysis führt damit zu unverzerrten Ergebnissen, wie beispielhaft die Schätzung $\hat{\rho}_{Y_1,Y_3}$ des Korrelationskoeffizienten von Y_1 und Y_3 in der Stichprobe zeigt: Sowohl aus den vollständig beobachteten als auch aus allen $n = 6$ Datensätzen ist $\hat{\rho}_{Y_1,Y_3} = -0{,}24$ zu schätzen, wodurch die Vernachlässigung der unvollständig beobachteten Untersuchungseinheiten in diesem Fall gerechtfertigt ist.

Ein weiteres, als Available Case Analysis bezeichnetes Eliminierungsverfahren berücksichtigt für die Datenauswertung die jeweils beobachteten Merkmale. Hierdurch tritt ein geringerer Informationsverlust als bei Anwendung der Complete Case Analysis auf, jedoch können sich inkonsistente Ergebnisse aufgrund der unterschiedlichen Stichprobenumfänge für verschiedene Statistiken ergeben.[34]

Fortsetzung von Beispiel 5.1:

Die Zufallsvariablen Y_1 und Y_3 sind vollständig beobachtet worden, so dass für die Schätzung des Korrelationskoeffizienten auf die Wertepaare (y_{i1}, y_{i3}) $(i = 1,\ldots,6)$ zurückgegriffen werden kann ($\hat{\rho}_{Y_1,Y_3} = -0{,}24$).

[34] Vgl. Schafer/Graham (2002), S. 155.

In der Praxis werden die Eliminierungsverfahren aufgrund ihrer einfach umzusetzenden Methodik häufig zur Behandlung von fehlenden Daten eingesetzt. Aufgrund der aufgeführten Nachteile ist deren Anwendung jedoch allenfalls bei Datenbeständen mit einem geringen Anteil an unvollständigen Datensätzen gerechtfertigt.

5.1.3 Mittelwertergänzung

Um ein vervollständigtes Datenmaterial zu erhalten, werden häufig in der Praxis die fehlenden Werte durch das arithmetische Mittel der beobachteten Werte einer Zufallsvariablen ersetzt. Im Folgenden soll gezeigt werden, dass diese Vorgehensweise oft zu verzerrten Schätzern und in vielen Fällen sogar zu schlechteren Ergebnissen als die Eliminierungsverfahren führt. Hierzu wird eine eindimensionale Zufallsvariable Y_1 in einer Grundgesamtheit vom Umfang N betrachtet. Weiterhin sind $y_{11},\ldots,y_{n1}$ die Realisationen einer einfachen Stichprobe vom Umfang n und s^2 bezeichne die zugehörige Stichprobenvarianz. Aus der Stichprobentheorie ist bekannt, dass der Stichprobenmittelwert $\bar{y}_1$ ein erwartungstreuer Schätzer für den Mittelwert von Y_1 in der Grundgesamtheit ist, der mit $\bar{Y}_1$ bezeichnet wird:[35]

$$E(\bar{y}_1) = \bar{Y}_1 \tag{5.1}$$

Bei einer einfachen Stichprobe gilt für die Varianz des Schätzers $\bar{y}_1$ [36]

$$\mathrm{Var}(\bar{y}_1) = \left(1 - \frac{n}{N}\right)\frac{\sigma^2}{n} \qquad (\mathrm{Var}(Y_1) = \sigma^2),$$

wobei σ^2 durch die Stichprobenvarianz s^2 geschätzt werden kann: [37]

$$\mathrm{Var}(\bar{y}_1) \approx \left(1 - \frac{n}{N}\right)\frac{s^2}{n} \tag{5.2}$$

Für einen hinreichend großen Stichprobenumfang n und für $n<<N$ ist $\bar{y}_1$ approximativ normalverteilt mit Erwartungswert $E(\bar{y}_1)$ und Varianz $\mathrm{Var}(\bar{y}_1)$:[38]

$$\bar{y}_1 \sim N\left(E(\bar{y}_1), \mathrm{Var}(\bar{y}_1)\right)$$

---

[35] Vgl. Cochran (1977), S. 22.
[36] Vgl. Cochran (1977), S. 26.
[37] Vgl. Cochran (1977), S. 26.
[38] Vgl. Cochran (1977), S. 39.

Aus (5.1) und (5.2) folgt

$$\bar{y}_1 \sim N\left(\overline{Y}_1, \frac{s^2}{n} - \frac{s^2}{N}\right),$$

so dass die Abweichung des Stichprobenmittelwerts $\bar{y}_1$ vom Mittelwert in der Grundgesamtheit $\overline{Y}_1$ ebenfalls approximativ normalverteilt ist mit Erwartungswert $E(\bar{y}_1 - \overline{Y}_1) = 0$ und Varianz $\text{Var}(\bar{y}_1 - \overline{Y}_1) = \frac{s^2}{n} - \frac{s^2}{N}$:[39]

$$\left(\bar{y}_1 - \overline{Y}_1\right) \sim N\left(0, \frac{s^2}{n} - \frac{s^2}{N}\right) \tag{5.3}$$

Ein 95%-Konfidenzintervall für die Zufallsgröße $\left(\bar{y}_1 - \overline{Y}_1\right)$ ist

$$\left[-1{,}96\, s^2\left(\frac{1}{n} - \frac{1}{N}\right); 1{,}96\, s^2\left(\frac{1}{n} - \frac{1}{N}\right)\right]. \tag{5.4}$$

Es seien im Folgenden lediglich *q* Werte in der Stichprobe vom Umfang *n* beobachtet worden. Der zugrunde liegende Ausfallmechanismus sei vom Typ MCAR, so dass die beobachteten Werte als eine einfache Stichprobe vom Umfang *q* aus der Grundgesamtheit betrachtet werden können. Demzufolge gelten für den Stichprobenmittelwert $\bar{y}_1^R$ der beobachteten Werte die selben Überlegungen wie in (5.3):[40]

$$\left(\bar{y}_1^R - \overline{Y}_1\right) \sim N\left(0, \frac{s_R^2}{q} - \frac{s_R^2}{N}\right)$$

95%-Konfidenzintervall zu $\left(\bar{y}_1^R - \overline{Y}_1\right)$:

$$\left[-1{,}96\, s_R^2\left(\frac{1}{q} - \frac{1}{N}\right); 1{,}96\, s_R^2\left(\frac{1}{q} - \frac{1}{N}\right)\right] \tag{5.5}$$

Dabei ist s_R^2 die Stichprobenvarianz der beobachteten Daten, welche erwartungstreu zum Schätzen der Varianz von Y_1 ist:

$$E(s_R^2) = \text{Var}(Y_1)$$

---

[39] Vgl. Rubin (1987), S. 13.
[40] Vgl. Rubin (1987), S. 13.

Werden die (*n*-*q*) fehlenden Werte durch den Stichprobenmittelwert der beobachteten Werte $\bar{y}_1^R$ ersetzt, erhält man den folgenden Ausdruck für die Stichprobenvarianz:

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{q} (y_{i1} - \bar{y}_1^R)^2 + \sum_{i=q+1}^{n} (y_{i1} - \bar{y}_1^R)^2 \right) = \frac{1}{n-1} \sum_{i=1}^{q} (y_{i1} - \bar{y}_1^R)^2 = \frac{q-1}{n-1} s_R^2$$

Die Stichprobenvarianz s^2 der erhobenen und ersetzten Werte ist nicht erwartungstreu zum Schätzen der Varianz von Y_1:

$$E(s^2) = E\left(\frac{q-1}{n-1} s_R^2 \right) = \frac{q-1}{n-1} E(s_R^2) \neq \mathrm{Var}(Y_1)$$

Durch die Mittelwertergänzung wird die Varianz von Y_1 um den Faktor $\frac{q-1}{n-1}$ unterschätzt. Eine anschließende Datenanalyse, welche den Ersetzungsprozess der Mittelwertergänzung nicht explizit berücksichtigt, führt nicht nur zu verzerrten Schätzern für die Varianz, sondern auch zu falschen Konfidenzintervallen. Im Beispiel ergibt sich – ohne Berücksichtigung der Imputation – für die ersetzten Daten das zu kleine Konfidenzintervall in (5.4). Dieser Bias wird – neben der bereits aufgeführten, geringeren Stichprobenvarianz – durch die Annahme eines zu großen Stichprobenumfangs *n* verstärkt, da die Stichprobengröße effektiv *q* ist. Dies ist anhand des Vergleichs von (5.4) mit dem wahren Konfidenzintervall aus (5.5) ersichtlich, da $s^2 < s_R^2$ und $q < n$ gilt. Durch diese methodischen Nachteile ist eine Anwendung der Mittelwertergänzung nur in seltenen Fällen gerechtfertigt.[41]

5.1.4 Gewichtungsverfahren

Eine weitere Methode zur Behandlung von fehlenden Werten besteht in der Gewichtung von vollständig beobachteten Datensätzen, um die damit verbundene Vernachlässigung der unvollständig beobachteten statistischen Einheiten zu kompensieren. Bei dem als Response Propensity Weighting bezeichneten Verfahren werden die Daten anhand der vollständig beobachteten Kovariaten zunächst in Klassen untergliedert. Anschließend wird jedem vollständig erhobenem Datensatz ein Gewicht

[41] Vgl. Rubin (1987), S. 14.

zugewiesen, welches der inversen Antwortwahrscheinlichkeit innerhalb der entsprechenden Klasse entspricht.[42]

Beispiel 5.2:

Gegeben sei die unvollständige Datenmatrix bestehend aus den Realisationen von zwei kardinal skalierten Zufallsvariablen Y_1 und Y_2:

i	Y_{i1}	Y_{i2}
1	3	4
2	2	6
3	3	?
4	4	?
5	4	1
6	4	?

i	R_{i1}	R_{i2}
1	0	0
2	0	0
3	0	1
4	0	1
5	0	0
6	0	1

Es ist die Antwortwahrscheinlichkeit für die jeweilige Klasse l zu bestimmen (l = 1, 2, 3), wobei die Klassenbildung anhand der vollständig beobachteten Zufallsvariable Y_1 erfolgt. Zu diesem Zweck wird bei der Anwendung des Gewichtungsverfahrens vorausgesetzt, dass für die bedingte Antwortwahrscheinlichkeit gegeben die Realisation y_1 gilt:

$$P_\psi(R_{i2} = 0 \mid Y_{i1} = 2) = \psi_1$$

$$P_\psi(R_{i2} = 0 \mid Y_{i1} = 3) = \psi_2$$

$$P_\psi(R_{i2} = 0 \mid Y_{i1} = 4) = \psi_3 \qquad (i = 1,...,n),\ \psi = (\psi_1, \psi_2, \psi_3)$$

Der Parametervektor $\psi = (\psi_1, \psi_2, \psi_3)$ ist aus den beobachteten Daten $\mathbf{y}_{obs}$ und der Indikatormatrix **r** zu schätzen:

$$\hat{\psi}_1 = 1 \quad , \quad \hat{\psi}_2 = 0{,}5 \quad , \quad \hat{\psi}_3 = 0{,}3\overline{3}$$

[42] Vgl. Little/Vartivarian (2003), S. 1589f.

Innerhalb des Gewichtungsverfahrens ist der Schätzer für den Mittelwert von Y_2 die gewichtete Summe der – aus den beobachteten Merkmalen bestimmten – Mittelwerte $\bar{y}_{2,l}^{R}$ in der jeweiligen Klasse l:

$$\bar{y}_2^{RPW} = \frac{1}{n}\sum_{l=1}^{3} \hat{\psi}_l^{-1} \bar{y}_{2,l}^{R} = 2{,}8\overline{3}$$

Der Schätzer in Beispiel 5.2 ist unverzerrt, wenn der Datenausfall nicht von den Werten der Zufallsvariable Y_2 abhängig und somit innerhalb der gebildeten Klassen zufällig ist. Diese Voraussetzung ist gleichbedeutend mit der Gültigkeit der MAR-Annahme in Kapitel 3, so dass – im Gegensatz zu den Eliminierungsverfahren – die Anwendung der Gewichtungsverfahren nicht auf zufälliges Fehlen der Werte im Sinne von MCAR begrenzt ist. Andererseits ist die Behandlung von fehlenden Werten mit Hilfe von Gewichtungsverfahren auf so genannte monotone Ausfallmuster beschränkt.[43] Diese besitzen für alle $i = 1,\ldots,n$ und $j = 1,\ldots,k\text{-}1$ die Eigenschaft, dass ein Wert y_{ij} immer dann beobachtet ist, wenn auch $y_{i(j+1)}$ bekannt ist. Formal gilt in diesem Fall für die Elemente r_{ij} einer realisierten Indikatormatrix **r**:

$$r_{ij} \leq r_{i(j+1)} \qquad \forall\, i = (1,\ldots,n),\, j = (1,\ldots,k-1) \tag{5.6}$$

Sind – wie im Beispiel 5.2 – die Realisationen von *einer* Zufallsvariable in der Stichprobe nicht beobachtet worden (univariater Datenausfall), so ist die Bedingung (5.6) stets erfüllt.

[43] Vgl. Little (1988), S. 294.

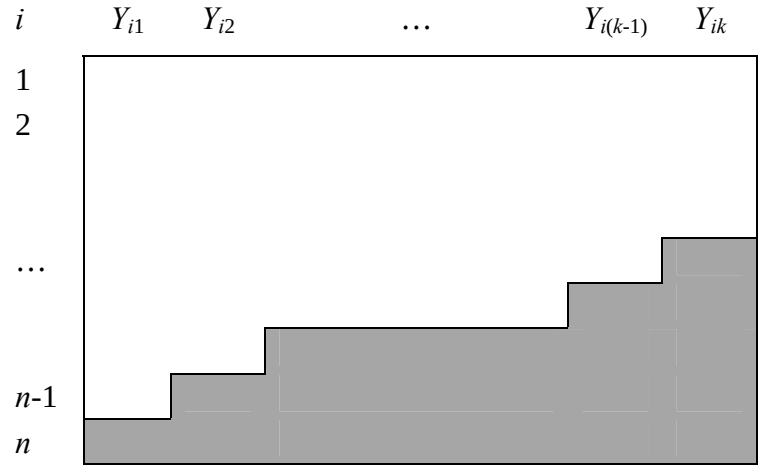


Abbildung 5.1: Allgemeine Darstellung eines monotonen Ausfallmusters (grau: fehlende Werte)

i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}
1	3	4	3	2	5
2	2	6	2	1	2
3	3	4	4	3	?
4	4	1	3	?	?
5	4	1	2	?	?
6	2	?	?	?	?

R_{i1}	R_{i2}	R_{i3}	R_{i4}	R_{i5}
0	0	0	0	0
0	0	0	0	0
0	0	0	0	1
0	0	0	1	1
0	0	0	1	1
0	1	1	1	1

Abbildung 5.2: Beispiel für ein monotones Muster des Datenausfalls

Ein häufiges Einsatzgebiet der Gewichtungsmethoden besteht in der Behandlung von vollständiger Antwortverweigerung (Unit Nonresponse), bei der unter Auswertung von zusätzlichen Informationen (Kovariaten) der Datenausfall erklärt und durch entsprechende Gewichtung berücksichtigt werden kann. Während die aus dem Gewichtungsverfahren resultierenden Schätzungen unter den genannten Voraussetzungen (Gültigkeit der MAR-Annahme, monotones Ausfallmuster) unverzerrt sind, können die aus den gewichteten Daten ermittelten Standardfehler eines Schätzers einen Bias aufweisen.[44]

[44] Vgl. Jones/Chromy (1982), S. 2.

Die Zweckmäßigkeit der Gewichtungsverfahren ist somit vom Analyseziel der anschließenden statistischen Auswertung abhängig, so dass der Anwendungsbereich dieser Verfahren auf die o.g. Fälle restringiert ist. Die in den nächsten Kapiteln beschriebenen Verfahren sind wesentlich flexibler und somit den traditionellen Methoden vorzuziehen.

5.2 Multiple Imputation von fehlenden Werten

5.2.1 Zielstellung des Verfahrens

Im Kontext der Behandlung von fehlenden Werten ist die Bereitstellung einer vervollständigten Datenmatrix von besonderer Bedeutung, um eine anschließende Datenanalyse mit statistischen Standardmethoden durchführen zu können. Bei der damit verbundenen Ersetzung von fehlenden Werten besteht das Problem, dass auch bei Bekanntheit oder Ignorierbarkeit des Ausfallmechanismus $P_{\psi}(\mathbf{R} \mid \mathbf{y})$ die ergänzten Werte von den wahren Ausprägungen der unbeobachteten Daten abweichen können. Diese Unsicherheit wird durch die Ersetzung von fehlenden Werten durch jeweils eine Ausprägung (single imputation) nicht berücksichtigt. Bei der anschließenden statistischen Analyse wird dann irrtümlich davon ausgegangen, dass die ergänzten Daten in gleicher Weise wie die tatsächlichen Beobachtungen $\mathbf{y}_{obs}$ erhoben wurden und somit bekannt sind.

Beispiel 5.3:

Es ist das arithmetische Mittel einer binären Zufallsvariable Y_2, die nur vereinzelt in einer Stichprobe vom Umfang $n = 10$ beobachtet wurde, zu schätzen. Eine weitere binäre Zufallsvariable Y_1 sei hingegen vollständig beobachtet worden.

	Y_{i1}	Y_{i2}
1	0	0
2	1	1
3	1	?
4	0	1
5	1	?
6	1	?
7	0	1
8	1	?
9	1	0
10	0	0

R_{i1}	R_{i2}
0	0
0	0
0	1
0	0
0	1
0	1
0	0
0	1
0	0
0	0

Es sei bekannt, dass der Ausfallmechanismus vom Typ MAR ist. Aufgrund dieser Information kann die bedingte Wahrscheinlichkeit von $Y_{i2} = 0$ gegeben $Y_{i1} = 1$ bei Nichtbeobachtung von Y_{i2} in der Stichprobe bestimmt werden:

$$P(Y_{i2} = 0 \mid Y_{i1} = 1, R_{i2} = 1) = P(Y_{i2} = 0 \mid Y_{i1} = 1, R_{i2} = 0) = 0{,}5 \quad (i = 1,...,10)$$

Durch das Ziehen von Werten aus der bekannten bedingten Verteilung $P(Y_{i2} \mid Y_{i1} = 1, R_{i2} = 1)$ können die Daten ergänzt werden. Bei dem so genannten Hot-Deck-Verfahren wird auf diese Weise jeweils ein fehlender Wert durch eine Ausprägung von Y_2 ersetzt.[45] Beispielhaft resultiert der folgende vervollständigte Datenbestand aus der Anwendung dieser single imputation Methode:

i	Y_{i1}	Y_{i2}
1	0	0
2	1	1
3	1	**0**
4	0	1
5	1	**0**
6	1	**1**
7	0	1
8	1	**1**
9	1	0
10	0	0

[45] Vgl. Huisman (1999), S. 96f.

Der Mittelwert in der Grundgesamtheit ist durch den Stichprobenmittelwert

$$\bar{y}_2 = \frac{1}{n}\sum_{i=1}^{n} y_{i2} = 0{,}5$$

zu schätzen. Die Varianz des Schätzers wird mit

$$s^2(\bar{y}_2) = \frac{1}{n}\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(y_{i2} - \bar{y}_2\right)^2\right) = 0{,}028$$

zu niedrig ausgewiesen, da diese Berechnung nur bei vollständiger Beobachtung aller Werte in der Stichprobe gültig ist und die fehlenden Werte lediglich geschätzt bzw. ergänzt wurden. Demzufolge ist

$$\left[\bar{y}_2 - t_{0,975}^{(n-1)}\, s(\bar{y}_2)\,;\, \bar{y}_2 + t_{0,975}^{(n-1)}\, s(\bar{y}_2)\right] = \left[0{,}12\,;\, 0{,}88\right]$$

auch kein 95%-Konfidenzintervall für den Mittelwert in der Grundgesamtheit, da man bei wiederholter Stichprobenziehung unter dem beschriebenen Ausfallmechanismus in weniger als 95% der Fälle ein Intervall erhält, welches den Mittelwert in der Grundgesamtheit beinhaltet.

Ein Ansatz zur Lösung dieser Problematik besteht in der mehrfachen Ergänzung von fehlenden Werten (multiple Imputation), bei dem mehrere plausibel vervollständigte Datenbestände erzeugt werden, die wiederum durch statistische Standardmethoden einzeln auswertbar sind. Wie im weiteren Verlauf der Arbeit gezeigt werden wird, sind die Parameterschätzer bei geeigneter Zusammenfassung der einzelnen Ergebnisse unverzerrt. Der wesentliche Vorteil des Verfahrens ist jedoch, dass die Verschiedenheit der Analyseergebnisse auf die Unsicherheit bezüglich der Ersetzung zurückzuführen ist und somit auch Aussagen über die Genauigkeit von Schätzern getroffen werden können.

5.2.2 Theoretische Grundlagen

Im Folgenden bezeichne $Q=Q(\mathbf{y})$ eine Größe (z.B. Mittelwert, Varianz, Korrelation, Regressionskoeffizient) in der Grundgesamtheit, die sich auf die Datenmatrix $\mathbf{y}$ bezieht und aus den beobachteten Daten zu schätzen ist. Ausgehend von der bayesiani-

schen Sichtweise besitzt Q eine Dichte, und die a-posteriori Verteilung von Q lässt sich formal durch

$$P(Q \mid \mathbf{y}_{obs}) = \int P(Q, \mathbf{y}_{mis} \mid \mathbf{y}_{obs})\, d\mathbf{y}_{mis}$$
$$= \int P(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) P(\mathbf{y}_{mis} \mid \mathbf{y}_{obs})\, d\mathbf{y}_{mis} \qquad (5.7)$$

darstellen.[46] Letztere ergibt sich somit aus allen möglichen a-posteriori Verteilungen von Q unter Vollständigkeit,[47] indem diese mit der Wahrscheinlichkeit für die entsprechenden Realisationen von $\mathbf{Y}_{mis}$ gewichtet werden. Gleichung (5.7) bildet die theoretische Grundlage der multiplen Imputation: Die interessierende a-posteriori Verteilung $P(Q \mid \mathbf{y}_{obs})$ kann durch Ziehen von m Werten aus $P(\mathbf{Y}_{mis} \mid \mathbf{y}_{obs})$ ermittelt werden, indem die resultierenden m a-posteriori Verteilungen $P(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis})$ geeignet zusammenzufassen sind.[48] Insbesondere gilt für den bedingten Erwartungswert von Q gegeben die beobachteten Daten:[49]

$$E(Q \mid \mathbf{y}_{obs}) = E\left[E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right] \qquad (5.8)$$

Dies impliziert, dass bei wiederholtem Ziehen von Werten für $\mathbf{Y}_{mis}$ die Erwartungswerte von Q unter Vollständigkeit $E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis})$ durchschnittlich dem Erwartungswert von Q unter Unvollständigkeit $E(Q \mid \mathbf{y}_{obs})$ entsprechen. Um den Erwartungswert unter Vollständigkeit $E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis})$ bestimmen zu können, wird die Verteilungsannahme

$$Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis} \sim N(\hat{Q}, U) \qquad (5.9)$$

bezüglich der a-posteriori Verteilung von Q getroffen, wobei $U = \mathrm{Var}(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis})$ und $\hat{Q}$ der Schätzer für Q unter Vollständigkeit ist.[50, 51] Dann gilt

$$E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) = \hat{Q} \qquad (5.10)$$

---

[46] Vgl. Rubin/Schenker (1986), S. 366.
[47] Wird davon ausgegangen, dass sowohl $\mathbf{y}_{obs}$ als auch $\mathbf{y}_{mis}$ beobachtet worden sind, so sei im Folgenden die a posteriori Verteilung von Q als a posteriori Verteilung unter Vollständigkeit bezeichnet.
[48] Die konkrete Anzahl m wird in den weiteren Ausführungen des Kapitels diskutiert.
[49] Vgl. Rubin/Schenker (1986); S. 367 sowie die Herleitung in Anhang A.1.
[50] Vgl. Rubin (1987), S. 75.
[51] Die Normalverteilungsannahme ist insbesondere bei einem großen Stichprobenumfang n gerechtfertigt.

und zusammen mit (5.8) folgt

$$E(Q \mid \mathbf{y}_{obs}) = E\big[E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\big] = E(\hat{Q} \mid \mathbf{y}_{obs}) . \tag{5.11}$$

Innerhalb der multiplen Imputation werden m ergänzte Datenmatrizen generiert und für jede dieser mit $\mathbf{y}^{(h)}$ $(1 \le h \le m)$ bezeichneten Matrizen wird der entsprechende Schätzer $\hat{Q}^{(h)}$ bestimmt. Zum Schätzen von $\hat{Q}$ ist das arithmetische Mittel $\frac{1}{m}\sum_{h=1}^{m}\hat{Q}^{(h)}$ geeignet, da

$$E(\hat{Q} \mid \mathbf{y}_{obs}) = E\left(\frac{1}{m}\sum_{h=1}^{m}\hat{Q}^{(h)} \mid \mathbf{y}_{obs}\right) \tag{5.12}$$

gilt und der Mittelwert die wirksamste Schätzfunktion für den Erwartungswert von $\hat{Q}$ ist. Aus den Formeln (5.11) und (5.12) erhält man

$$E(Q \mid \mathbf{y}_{obs}) = E\left(\frac{1}{m}\sum_{h=1}^{m}\hat{Q}^{(h)} \mid \mathbf{y}_{obs}\right), \tag{5.13}$$

so dass die interessierende Größe Q aus den Einzelergebnissen $\hat{Q}^{(1)},...,\hat{Q}^{(m)}$ unverzerrt durch

$$\overline{\hat{Q}} = \frac{1}{m}\sum_{h=1}^{m}\hat{Q}^{(h)} \tag{5.14}$$

geschätzt werden kann. Folglich können unter der Voraussetzung, dass die Verteilung $P(\mathbf{Y}_{mis} \mid \mathbf{y}_{obs})$ korrekt spezifiziert wurde, statistische Standardverfahren zur separaten Berechnung von $\hat{Q}^{(1)},...,\hat{Q}^{(m)}$ verwendet werden, deren Ergebnisse durch anschließende Mittelwertbildung zusammenzufassen sind.

Ein genereller Vorteil der Imputation von fehlenden Werten ist die strikte Trennung vom Ersetzungsprozess und anschließender Datenauswertung. Insofern ist die Ersetzung als eine Datenvorverarbeitung zu betrachten, und es muss gesichert sein, dass die Berechnung jeglicher Statistiken aus den vervollständigten Datenbeständen zu validen Schlussfolgerungen bezüglich der Grundgesamtheit führt. So ist es z.B. für die Bestimmung von Konfidenzintervallen für Q notwendig, dass neben dem Schätzer für Q auch die aus den Daten ermittelte Varianz des Schätzers unverzerrt ist. Der

Beweis, dass auch die letztgenannte Forderung durch die multiple Imputation erfüllt ist, führt über die allgemein gültige Zerlegung der Varianz von Q:[52]

$$\mathrm{Var}(Q \mid \mathbf{y}_{obs}) = E\left[\mathrm{Var}(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right] + \mathrm{Var}\left[E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right] \tag{5.15}$$

Aufgrund der Annahme (5.9) gilt für den ersten Summanden in (5.15) per Definition von U bei multipler Imputation:

$$E\left[\mathrm{Var}(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right] = E(U \mid \mathbf{y}_{obs}) \tag{5.16}$$

Die mit $U^{(h)}$ bezeichnete Varianz von Q kann bei Bekanntheit der Datenmatrix $\mathbf{y}^{(h)}$ für $h = 1,...,m$ berechnet werden, und das arithmetische Mittel dieser m Werte

$$\overline{U} = \frac{1}{m}\sum_{h=1}^{m} U^{(h)} \tag{5.17}$$

ist der Schätzer für den Erwartungswert von U, weil

$$E(U \mid \mathbf{y}_{obs}) = E\left(\frac{1}{m}\sum_{h=1}^{m} U^{(h)} \mid \mathbf{y}_{obs}\right) \tag{5.18}$$

gilt.

Die Zufallsvariable $\overline{U}$ wird von Rubin/Schenker (1986) als „Average Within-Imputation Variance“ von Q bezeichnet, da sie den durchschnittlichen Schätzfehler in einer einzelnen vervollständigten Matrix repräsentiert.[53]

Für den zweiten Summanden in (5.15) gilt wegen (5.10)

$$\mathrm{Var}\left[E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right] = \mathrm{Var}\left(\hat{Q} \mid \mathbf{y}_{obs}\right), \tag{5.19}$$

und die Varianz von $\hat{Q}$ ist durch die Stichprobenvarianz

$$B = \frac{1}{m-1}\sum_{h=1}^{m}\left(\hat{Q}^{(h)} - \overline{\hat{Q}}\right)^2 \tag{5.20}$$

als deren wirksamste Schätzfunktion approximativ zu bestimmen:

$$\mathrm{Var}\left(\hat{Q} \mid \mathbf{y}_{obs}\right) = E\left(\frac{1}{m-1}\sum_{h=1}^{m}\left(\hat{Q}^{(h)} - \overline{\hat{Q}}\right)^2 \mid \mathbf{y}_{obs}\right) \tag{5.21}$$

---

[52] Vgl. Rubin/Schenker (1986), S. 367 sowie die Herleitung in Anhang A.2.
[53] Vgl. Rubin/Schenker (1986), S. 367.

Folglich bemisst *B* die Schwankung von $\hat{Q}$, die aus den *m* Schätzungen für $\hat{Q}$ resultiert. Letztendlich sind diese Abweichungen allein auf die unterschiedlichen Ersetzungen in den *m* Matrizen zurückzuführen, und somit spiegelt *B* die Unsicherheit bezüglich der fehlenden Werte wider. Treffenderweise definieren Rubin/Schenker (1986) diese Zufallsvariable auch als „Between-Imputation Variance" von $\hat{Q}$.[54]

Wie gezeigt wurde, ist der quadrierte Standardfehler von *Q* in (5.15) die Summe aus der „Average Within-Imputation Variance" $\overline{U}$ und der „Between-Imputation Variance" *B*. Da die letztgenannte Varianz bei einer begrenzten Anzahl an multiplen Imputationen je fehlendem Wert zu gering ausgewiesen wird,[55] ist der Korrekturfaktor

$$1+\frac{1}{m}$$

zu verwenden,[56] und die Varianz von *Q* ist durch

$$T=\overline{U}+\left(1+\frac{1}{m}\right)B \tag{5.22}$$

zu schätzen.

Enthalten die fehlenden Werte keine Information über die untersuchte Größe *Q*, so sind die Schätzungen $\hat{Q}^{(1)},...,\hat{Q}^{(m)}$ identisch und die Varianz *B*, die aus den Imputationen resultiert, ist Null. In diesem Fall ist die Schätzung der Varianz von *Q* durch $\overline{U}$ ausreichend. Ausgehend von dieser Überlegung wird eine Rate

$$c=\frac{\left(1+\frac{1}{m}\right)B}{\overline{U}} \tag{5.23}$$

definiert, welche den relativen Anstieg der Varianz von *Q* aufgrund des Datenausfalls misst.[57] Dementsprechend kann der Anteil der aus der Nichtbeobachtung resultierenden Varianz an der gesamten Varianz von *Q* durch

---

[54] Vgl. Rubin/Schenker (1986), S. 367.
[55] Vgl. Rubin (1987), S. 87.
[56] Vgl. zur Herleitung des Korrekturfaktors Rubin (1987), S. 87ff.
[57] Vgl. Rubin (1987), S. 77.

$$\tau = \frac{\left(1+\frac{1}{m}\right)B}{\left(1+\frac{1}{m}\right)B+\overline{U}} = \frac{c}{1+c} \tag{5.24}$$

dargestellt werden.[58]

Die Schätzer für Q und dessen Varianz sind bei einer begrenzten Anzahl von m Ersetzungen nicht effizient. Dies hat zur Konsequenz, dass z.B. ein Konfidenzintervall für Q mit einer geringeren Länge angegeben werden könnte, wenn m wesentlich größer gewählt wird. Wie Rubin (1987) zeigte, erhöht sich der Schätzfehler bei m Imputationen im Vergleich zu unendlich vielen Ergänzungen näherungsweise um den Faktor

$$\sqrt{1+\frac{\tau}{m}}.$$[59]

Ist beispielsweise die Rate $c = 1$, so wird bei $m = 5$ Imputationen der Standardfehler lediglich um 5% höher als im Fall von $m = \infty$ ausgewiesen.[60] Selbst bei einem hohen Anteil fehlender Informationen ist somit eine geringe Anzahl an Ersetzungen in den meisten Anwendungen ausreichend, da der Schätzfehler nur unwesentlich größer als bei hoher Anzahl m ist. In der Literatur werden dementsprechend 3-10 Imputationen je fehlendem Wert als ausreichend betrachtet.[61]

Insbesondere aus dem Grund der Realisierbarkeit ist es bei großen Datenbeständen sogar notwendig, die Anzahl an Ersetzungen m stark zu begrenzen. Würde hingegen m sehr hoch gewählt werden können, so ist die Normalverteilungsannahme für den Schätzfehler

$$\left(Q-\overline{\hat{Q}}\right) \sim N(0,T) \tag{5.25}$$

plausibel,[62] und

[58] Vgl. Schafer (1999), S. 5.
[59] Vgl. Rubin (1987), S. 114.
[60] Vgl. Schafer (1997), S. 107. Die Berechnung basiert auf dem o.g. Faktor.
[61] Vgl. Rubin (1987), S. 114 und Schafer (1997), S. 106f.
[62] Vgl. Rubin (1996), S. 477.

$$\left[\overline{\hat{Q}} - z_{1-\frac{\alpha}{2}} \sqrt{T} \; ; \overline{\hat{Q}} + z_{1-\frac{\alpha}{2}} \sqrt{T} \right] \tag{5.26}$$

ist ein Konfidenzintervall für Q. Werden andererseits – wie in der Praxis üblich – nur wenige Ersetzungen je fehlendem Wert durchgeführt, ist vielmehr von

$$\frac{Q - \hat{Q}}{\sqrt{T}} \sim t^{df}_{1-\frac{\alpha}{2}} \tag{5.27}$$

als Verteilung für den standardisierten Schätzfehler auszugehen.[63] Die Anzahl der Freiheitsgrade *df* der t-Verteilung ist dabei durch

$$df = (m-1)\left(1 + \frac{1}{c}\right)^2 \tag{5.28}$$

mit der Rate c aus (5.23) festzulegen.[64] Sowohl für $m \to \infty$ als auch für $c \to 0$ entspricht die t-Verteilung der Standardnormalverteilung. Der Fall $c \to 0$ verdeutlicht erneut, dass bei geringer „Between Imputation Variance“ B zahlreiche Ersetzungen je fehlendem Wert zu keiner deutlich höheren Effizienz der Schätzungen führen, da die m Imputationen je fehlendem Wert nicht wesentlich voneinander abweichen.

Fortsetzung von Beispiel 5.3:

Die multiple Imputation eines fehlenden Wertes aus Beispiel 5.3 soll durch dreimaliges Ziehen ($m = 3$) aus der bekannten bedingten Verteilung $P(Y_{i2} \mid Y_{i1} = 1, R_{i2} = 1)$ ($i = 1,\ldots,n$) erfolgen. Die folgenden Datenmatrizen $\mathbf{y}^{(1)}$, $\mathbf{y}^{(2)}$ und $\mathbf{y}^{(3)}$ stellen ein mögliches Ergebnis dieser Vorgehensweise dar.

[63] Vgl. Schafer (1999), S. 4.
[64] Vgl. Rubin (1987), S. 77.

i	$Y_{i1}^{(1)}$	$Y_{i2}^{(1)}$
1	0	0
2	1	1
3	1	**1**
4	0	1
5	1	**1**
6	1	**1**
7	0	1
8	1	**0**
9	1	0
10	0	0

i	$Y_{i1}^{(2)}$	$Y_{i2}^{(2)}$
1	0	0
2	1	1
3	1	**1**
4	0	1
5	1	**1**
6	1	**0**
7	0	1
8	1	**0**
9	1	0
10	0	0

i	$Y_{i1}^{(3)}$	$Y_{i2}^{(3)}$
1	0	0
2	1	1
3	1	**0**
4	0	1
5	1	**0**
6	1	**1**
7	0	1
8	1	**0**
9	1	0
10	0	0

Das arithmetische Mittel der Stichprobenmittelwerte in den drei Datenbeständen $\bar{y}_2^{(1)}, \bar{y}_2^{(2)}$ und $\bar{y}_2^{(3)}$ ist der Schätzer für den Mittelwert von Y_2 in der Grundgesamtheit:

$$\bar{y}_2 = \frac{1}{m}\sum_{h=1}^{m}\bar{y}_2^{(h)} = \frac{1}{m}\sum_{h=1}^{m}\left(\frac{1}{n}\sum_{i=1}^{n}y_{i2}^{(h)}\right) = 0{,}5$$

Die mit T bezeichnete Varianz des Schätzers setzt sich additiv aus dem durchschnittlichen Schätzfehler

$$\overline{U} = \frac{1}{m}\sum_{h=1}^{m}\left(s_{\bar{y}_2}^{(h)}\right)^2 = \frac{1}{m}\sum_{h=1}^{m}\left(\frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(y_{i2}^{(h)} - \bar{y}_2^{(h)})^2\right) = 0{,}027$$

und der Varianz aufgrund der Imputationen

$$B = \frac{1}{m-1}\sum_{h=1}^{m}\left(\bar{y}_2^{(h)} - \bar{y}_2\right)^2 = 0{,}01$$

zusammen:

$$T = \overline{U} + \left(1 + \frac{1}{m}\right)B = 0{,}04$$

Unter der Prämisse, dass die MAR-Annahme erfüllt ist und somit der Ausfallmechanismus korrekt spezifiziert wurde, ist aufgrund von (5.26) der Bereich

$$\left[\bar{y}_2 - t_{0{,}975}^{df}\sqrt{T}\ ;\ \bar{y}_2 + t_{0{,}975}^{df}\sqrt{T}\right] = \left[0{,}08\ ;\ 0{,}92\right]$$

ein 95%-Konfidenzintervall für den Mittelwert in der Grundgesamtheit.

Obwohl 40% der Werte von Y_2 in der Stichprobe fehlen und der Anteil der Varianz aufgrund der Nichtbeobachtung an der gesamten Varianz mit $\tau = 0{,}33$ relativ hoch ist, ist das Intervall in Beispiel 5.3 nur geringfügig kleiner als das tatsächliche Konfidenzintervall. Der Grund hierfür besteht in dem geringen Stichprobenumfang von $n = 10$, wodurch – wie im Beispiel 5.3 unterstellt – ein Intervall bei vollständiger Beobachtung von Y_2 in der Stichprobe bereits entsprechend groß ist. Werden hingegen umfangreiche Datenbestände statistisch ausgewertet, bewirkt die Berücksichtigung der Unsicherheit, die sich aus dem Datenausfall ergibt, im Allgemeinen deutlichere Unterschiede bei der Varianz des Schätzers und damit wesentlich abweichende Unter- und Obergrenzen des Intervalls.

5.3 Likelihood-basierte Verfahren

In der statistischen Literatur zur Behandlung von fehlenden Werten werden zunehmend Lösungsansätze diskutiert, die auf der Bestimmung des Maximum-Likelihood-Schätzers für den unbekannten Parameter θ der Wahrscheinlichkeitsfunktion $P_\theta(\mathbf{y})$ beruhen. Diese likelihood-basierten Verfahren sind aufgrund von (4.3) insbesondere in den Fällen anwendbar, in denen der zugrunde liegende Ausfallmechanismus ignorierbar ist. Ein Vorteil der Methoden besteht, im Gegensatz zu den meisten traditionellen Verfahren, in der Einbeziehung *aller* verfügbaren Informationen des Datenbestandes in Form von $\mathbf{y}_{obs}$ für die Ergänzung fehlender Werte. Weiterhin sind likelihood-basierte Verfahren unabhängig von der Erscheinungsform der fehlenden Werte in der Datenmatrix $\mathbf{y}$ einsetzbar. Hingegen sind traditionelle Methoden häufig auf ein monotones Ausfallmuster beschränkt bzw. auf die Erfüllung der restriktiven MCAR-Annahme angewiesen.

Die Likelihood-Funktion $L(\mathbf{y}_{obs} \mid \theta)$ ist im Allgemeinen kompliziert und nicht in geschlossener Form darstellbar, so dass die Bestimmung des Schätzers von θ auf die herkömmliche Weise nicht möglich ist.[65] Das Maximierungsproblem kann in diesen Fällen durch die Anwendung von Näherungsverfahren, wie z.B. dem Expectation-Maximization-Algorithmus (EM-Algorithmus), gelöst werden.

[65] Vgl. Schafer (1997), S. 37.

5.3.1 EM-Algorithmus

Der EM-Algorithmus ist ein iteratives Verfahren, bei dem unter Verwendung eines vorläufigen Parameterschätzers die fehlenden Werte ergänzt werden und daran anschließend für die gesamten, vervollständigten Daten ein neuer Schätzer des Parametervektors bestimmt wird.[66]

Für die Loglikelihood-Funktion $l(\mathbf{y}_{obs} \mid \theta)$ der beobachteten Daten

$$\mathbf{y}_{obs} = (\underline{y}_{obs,1},...,\underline{y}_{obs,n})^T$$

gilt aufgrund der Unabhängigkeit der Untersuchungseinheiten:[67]

$$l(\mathbf{y}_{obs}|\theta) = \sum_{i=1}^{n} \ln P_\theta(\underline{y}_{obs,i}) = \sum_{i=1}^{n} \ln \frac{P_\theta(\underline{y}_{obs,i}, \underline{y}_{mis,i})}{P_\theta(\underline{y}_{mis,i} \mid \underline{y}_{obs,i})}$$

$$= \sum_{i=1}^{n} \ln P_\theta(\underline{y}_{obs,i}, \underline{y}_{mis,i}) - \sum_{i=1}^{n} \ln P_\theta(\underline{y}_{mis,i} \mid \underline{y}_{obs,i})$$

$$= l(\mathbf{y} \mid \theta) - \sum_{i=1}^{n} \ln P_\theta(\underline{y}_{mis,i} \mid \underline{y}_{obs,i}) \qquad (5.29)$$

Die Loglikelihood-Funktion der unvollständig beobachteten Daten entspricht nach (5.29) der Loglikelihood der gesamten Daten abzüglich der prädiktiven bedingten Verteilung von den fehlenden Werten gegeben die beobachteten Realisationen. Es kann gezeigt werden, dass ein Anstieg der vollständigen Loglikelihood ebenfalls eine Erhöhung der unvollständigen Loglikelihood zur Folge hat.[68] Hierdurch kann das komplizierte Maximierungsproblem von θ bei der unvollständigen Loglikelihood durch die Ermittlung des leichter zu bestimmenden ML-Schätzers von θ unter Vollständigkeit gelöst werden.

Im ersten E-Schritt („Expectation-Step“) des Algorithmus wird der bedingte Erwartungswert der vollständigen Loglikelihood gegeben die beobachteten Daten und un-

[66] Vgl. Schafer/Graham (2002), S. 163.
[67] Vgl. Rässler (2000), S. 73.
[68] Vgl. Dempster et al. (1977), S. 6f.

ter Festlegung eines Startwertes $\theta^{(0)}$ gebildet. Der Erwartungswert ist dabei von θ abhängig und kann als eine Funktion $V(\theta \mid \theta^{(0)})$ definiert werden:[69]

$$V(\theta \mid \theta^{(0)}) = E[l(\mathbf{y} \mid \theta) \mid \mathbf{y}_{obs}, \theta^{(0)}] = \int l(\mathbf{y} \mid \theta) P_{\theta}(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \theta^{(0)}) d\mathbf{y}_{mis} \tag{5.30}$$

Die bedingte Verteilung $P_{\theta}(\mathbf{Y}_{mis} \mid \mathbf{y}_{obs}, \theta^{(0)})$ verdeutlicht hierbei den Zusammenhang zwischen den fehlenden Werten und dem unbekannten Parametervektor. Durch die Integration über $\mathbf{y}_{mis}$ in (5.30) ist der Expectation-Step mit dem Ersetzen der fehlenden Daten durch deren bedingte Erwartungswerte vergleichbar.[70]

Im folgenden M-Schritt („Maximization-Step“) ist der Parameterwert $\theta^{(1)}$ so zu wählen, dass $V(\theta \mid \theta^{(0)})$ maximiert wird:[71]

$$V(\theta^{(1)} \mid \theta^{(0)}) \geq V(\theta \mid \theta^{(0)}) \text{ für alle } \theta \in \mathbf{\Omega}_{\theta} \tag{5.31}$$

Im weiteren Ablauf des Algorithmus wird der E-Schritt mit $\theta^{(1)}$ als neuem Parameterwert erneut ausgeführt und $V(\theta \mid \theta^{(1)})$ im anschließenden M-Schritt maximiert. Dieses iterative Vorgehen wird wiederholt, bis sich die Parameterschätzungen von zwei aufeinander folgenden Iterationen *t* und *t*+1 nicht mehr wesentlich unterscheiden bzw. deren Differenz eine im Vorhinein festgelegte Grenze ε nicht mehr überschreitet:

$$\left|\theta^{(t+1)} - \theta^{(t)}\right| \leq \varepsilon$$

Bei der Verwendung dieses Abbruchkriteriums ist nicht auszuschließen, dass der Algorithmus gegen ein lokales Maximum konvergiert. Um dieses Problem zu umgehen, wird empfohlen, den EM-Algorithmus zur Bestimmung eines globalen Maximums mehrmals mit unterschiedlichen Startwerten $\theta^{(0)}$ durchzuführen.[72]

---

[69] Vgl. Little/Rubin (2002), S. 168.
[70] Vgl. Schafer (1997), S. 39.
[71] Vgl. Little/Rubin (2002), S. 168.
[72] Vgl. Schafer (1997), S. 52.

Beispiel 5.4:

Es werden n Untersuchungseinheiten von zwei dichotomen Variablen Y_1 und Y_2 betrachtet, wobei n_{ab} $(a,b=0,1)$ die Anzahl der Untersuchungseinheiten mit $Y_1 = a$ und $Y_2 = b$ bezeichnet. Weiterhin ist p_{ab} die Wahrscheinlichkeit, dass die Zufallsvariablen Y_1 und Y_2 die Werte a bzw. b annehmen. Unter Beachtung der iid-Eigenschaft der Stichprobe sind die absoluten Häufigkeiten (n_{00}, n_{01}, n_{10}, n_{11}) Realisationen der mehrdimensionalen Zufallsvariable (N_{00}, N_{01}, N_{10}, N_{11}), die multinomialverteilt ist mit den Parametern n und $\theta = (p_{00}, p_{01}, p_{10}, p_{11})$: [73]

$$P_\theta(N_{00}=n_{00},...,N_{11}=n_{11}) = \frac{n!}{n_{00}!n_{01}!n_{10}!n_{11}!}(p_{00})^{n_{00}}(p_{01})^{n_{01}}(p_{10})^{n_{10}}(p_{11})^{n_{11}}$$

$$(n = n_{00}+n_{01}+n_{10}+n_{11}) \qquad (5.32)$$

Ferner beschreibt n_{ab}^R die absolute Häufigkeit der vollständig beobachteten Untersuchungseinheiten mit $Y_1 = a$ und $Y_2 = b$ und

$$n^R = n_{00}^R + n_{01}^R + n_{10}^R + n_{11}^R$$

ist die gesamte Anzahl der vollständigen Datensätze.

Die absolute Häufigkeit der Untersuchungseinheiten mit $Y_1 = a$ und fehlendem Wert der Zufallsvariable Y_2 wird im Folgenden mit $n_{a\bullet}^{Y_1+}$ bezeichnet, während $n_{\bullet b}^{+Y_2}$ der Anzahl der Datensätze mit $Y_2 = b$ und fehlender Ausprägung von Y_1 entspricht. Weiterhin sind die Randhäufigkeiten in den vollständigen und unvollständigen Datensätzen durch

$$n_{\bullet b}^R = n_{0b}^R + n_{1b}^R \quad n_{a\bullet}^R = n_{a0}^R + n_{a1}^R \quad n^{Y_1+} = n_{0\bullet}^{Y_1+} + n_{1\bullet}^{Y_1+} \quad n^{+Y_2} = n_{\bullet 0}^{+Y_2} + n_{\bullet 1}^{+Y_2} \quad \forall a,b = 0,1$$

definiert.

Auf dieser Einteilung basierend beinhaltet Abbildung 5.3 neben der Kontingenztafel für die vollständig beobachteten Merkmalsträger die bekannten Randhäufigkeiten bei Datenausfall von Y_1 oder Y_2.

---

[73] Vgl. Schafer (1997), S. 42.

Y_1 und Y_2 beobachtet

	Y_2=0	Y_2=1	
Y_1=0	$n_{00}^{R}=5$	$n_{01}^{R}=15$	$n_{0\bullet}^{R}=20$
Y_1=1	$n_{10}^{R}=20$	$n_{11}^{R}=10$	$n_{1\bullet}^{R}=30$
	$n_{\bullet 0}^{R}=25$	$n_{\bullet 1}^{R}=25$	$n^{R}=50$

Y_2 nicht beobachtet

	$Y_2\in\{0;1\}$
Y_1=0	$n_{0\bullet}^{Y_1+}=15$
Y_1=1	$n_{1\bullet}^{Y_1+}=15$
	$n^{Y_1+}=30$

Y_1 nicht beobachtet

$Y_1\in\{0;1\}$	Y_2=0	Y_2=1	
	$n_{\bullet 0}^{+Y_2}=15$	$n_{\bullet 1}^{+Y_2}=5$	$n^{+Y_2}=20$

Abbildung 5.3: Darstellung der beobachteten, absoluten (Rand-)Häufigkeiten von zwei dichotomen Zufallsvariablen

Bei vollständiger Beobachtung der Daten kann die gemeinsame Verteilung der Zufallsvariablen Y_1 und Y_2 durch die Parameter der Multinomialverteilung in (5.32) beschrieben werden. Die zugehörige Loglikelihood-Funktion ist in der folgenden Weise darstellbar:[74]

$$l(n_{00},...,n_{11}\mid\theta)=\log\left(\frac{n!}{n_{00}!n_{01}!n_{10}!n_{11}!}\right)+\sum_{a=0}^{1}\sum_{b=0}^{1}n_{ab}\log(p_{ab})$$

In der ersten Iteration des EM-Algorithmus ist der Erwartungswert der Loglikelihood-Funktion gegeben die beobachteten Daten

$$n^{obs}=\{n_{00}^{R},n_{01}^{R},n_{10}^{R},n_{11}^{R},n_{0\bullet}^{Y_1+},n_{1\bullet}^{Y_1+},n_{\bullet 0}^{+Y_2},n_{\bullet 1}^{+Y_2}\}$$

und eines Startwertes

$$\theta^{(0)}=(\hat{p}_{00}^{(0)},\hat{p}_{01}^{(0)},\hat{p}_{10}^{(0)},\hat{p}_{11}^{(0)})$$

zu bilden:

$$E\left[l(n_{00},...,n_{11}\mid\theta)\mid n^{obs},\theta^{(0)}\right]=E\left[\left(\log\left(\frac{n!}{n_{00}!n_{01}!n_{10}!n_{11}!}\right)+\sum_{a=0}^{1}\sum_{b=0}^{1}n_{ab}\log(p_{ab})\right)\mid n^{obs},\theta^{(0)}\right]$$

$$=\log\left(\frac{n!}{n_{00}!n_{01}!n_{10}!n_{11}!}\right)+\sum_{a=0}^{1}\sum_{b=0}^{1}E(n_{ab}\mid n^{obs},\theta^{(0)})\log(p_{ab}) \qquad (5.33)$$

Eine mögliche Ausgangslösung für den Vektor $\theta^{(0)}$ besteht aus den relativen Häufigkeiten von Merkmalskombinationen in den vollständig beobachteten Datensätzen:

[74] Vgl. Schafer (1997), S. 42.

$$\hat{p}_{ab}^{(0)} = \frac{n_{ab}^{R}}{n^{R}} \qquad \forall\ a,b = 0,1$$

Im Beispiel erhält man die folgenden Elemente des Parametervektors $\theta^{(0)}$:

$$\hat{p}_{00}^{(0)} = \frac{n_{00}^{R}}{n^{R}} = 0{,}1 \quad \hat{p}_{01}^{(0)} = \frac{n_{01}^{R}}{n^{R}} = 0{,}3 \quad \hat{p}_{10}^{(0)} = \frac{n_{10}^{R}}{n^{R}} = 0{,}4 \quad \hat{p}_{11}^{(0)} = \frac{n_{11}^{R}}{n^{R}} = 0{,}2$$

Die Anzahl der Datensätze mit $Y_1 = a$ und $Y_2 = b$ setzt sich aus den korrespondierenden, vollständig beobachteten Untersuchungseinheiten n_{ab}^{R} sowie den Datensätzen mit einer nicht beobachteten Variable ($n_{ab}^{Y_1+}$ und $n_{ab}^{+Y_2}$) zusammen:

$$n_{ab} = n_{ab}^{R} + n_{ab}^{Y_1+} + n_{ab}^{+Y_2} \qquad \forall \quad a,b = 0,1$$

Die zu schätzende Anzahl der Untersuchungseinheiten mit einer fehlenden Variable ist – gegeben die Randhäufigkeit $n_{a\bullet}^{Y_1+}$ bzw. $n_{\bullet b}^{+Y_2}$ – ebenfalls multinomialverteilt mit den Parametern $\theta = \left(\frac{p_{a0}}{p_{a\bullet}}, \frac{p_{a1}}{p_{a\bullet}} \right)$ bzw. $\theta = \left(\frac{p_{0b}}{p_{\bullet b}}, \frac{p_{1b}}{p_{\bullet b}} \right)$. Im ersten Schritt des EM-Algorithmus wird somit – unter Berücksichtigung des Startwertes $\theta^{(0)}$ – von den folgenden Verteilungen der fehlenden absoluten Häufigkeiten ausgegangen:[75]

$$\left(n_{a0}^{Y_1+}, n_{a1}^{Y_1+} \right) | \, n^{obs}, \theta^{(0)} \sim M\left(n_{a\bullet}^{Y_1+}, \frac{\hat{p}_{a0}^{(0)}}{\hat{p}_{a\bullet}^{(0)}}, \frac{\hat{p}_{a1}^{(0)}}{\hat{p}_{a\bullet}^{(0)}} \right) \quad \forall\ a = 0,1$$

$$\left(n_{0b}^{+Y_2}, n_{1b}^{+Y_2} \right) | \, n^{obs}, \theta^{(0)} \sim M\left(n_{\bullet b}^{+Y_2}, \frac{\hat{p}_{0b}^{(0)}}{\hat{p}_{\bullet b}^{(0)}}, \frac{\hat{p}_{1b}^{(0)}}{\hat{p}_{\bullet b}^{(0)}} \right) \quad \forall\ b = 0,1$$

Im Beispiel sind die absoluten, nicht beobachteten Zellhäufigkeiten wie folgt verteilt:

$$\left(n_{00}^{Y_1+}, n_{01}^{Y_1+} \right) | \, n^{obs}, \theta^{(0)} \sim M\left(15, \frac{0{,}1}{0{,}4}, \frac{0{,}3}{0{,}4} \right)$$

$$\left(n_{10}^{Y_1+}, n_{11}^{Y_1+} \right) | \, n^{obs}, \theta^{(0)} \sim M\left(15, \frac{0{,}4}{0{,}6}, \frac{0{,}1}{0{,}6} \right)$$

$$\left(n_{00}^{+Y_2}, n_{10}^{+Y_2} \right) | \, n^{obs}, \theta^{(0)} \sim M\left(15, \frac{0{,}1}{0{,}5}, \frac{0{,}4}{0{,}5} \right)$$

[75] Vgl. Schafer (1997), S. 44.

$$\left(n_{01}^{+Y_2}, n_{11}^{+Y_2}\right) | \, n^{obs}, \theta^{(0)} \sim M\left(5, \frac{0{,}3}{0{,}5}, \frac{0{,}2}{0{,}5}\right)$$

Im E-Schritt des Algorithmus ergibt sich

$$E(n_{ab} \mid n^{obs}, \theta^{(0)}) = E(n_{ab}^R + n_{ab}^{Y_1+} + n_{ab}^{+Y_2} \mid n^{obs}, \theta^{(0)})$$

$$= n_{ab}^R + n_{a\bullet}^{Y_1+} \frac{\hat{p}_{ab}^{(0)}}{\hat{p}_{a\bullet}^{(0)}} + n_{\bullet b}^{+Y_2} \frac{\hat{p}_{ab}^{(0)}}{\hat{p}_{\bullet b}^{(0)}}$$

für den Erwartungswert der absoluten Zellhäufigkeiten in Formel (5.33).[76] Die auf diese Weise geschätzten Häufigkeiten sind im Beispiel

$$E(n_{00} \mid n^{obs}, \theta^{(0)}) = n_{00}^R + n_{0\bullet}^{Y_1+} \frac{\hat{p}_{00}^{(0)}}{\hat{p}_{0\bullet}^{(0)}} + n_{\bullet 0}^{+Y_2} \frac{\hat{p}_{00}^{(0)}}{\hat{p}_{\bullet 0}^{(0)}} = 5 + 15\frac{0{,}1}{0{,}4} + 15\frac{0{,}1}{0{,}5} = 11{,}75$$

$$E(n_{01} \mid n^{obs}, \theta^{(0)}) = n_{01}^R + n_{0\bullet}^{Y_1+} \frac{\hat{p}_{01}^{(0)}}{\hat{p}_{0\bullet}^{(0)}} + n_{\bullet 1}^{+Y_2} \frac{\hat{p}_{01}^{(0)}}{\hat{p}_{\bullet 1}^{(0)}} = 15 + 15\frac{0{,}3}{0{,}4} + 5\frac{0{,}3}{0{,}5} = 29{,}25$$

$$E(n_{10} \mid n^{obs}, \theta^{(0)}) = n_{10}^R + n_{1\bullet}^{Y_1+} \frac{\hat{p}_{10}^{(0)}}{\hat{p}_{1\bullet}^{(0)}} + n_{\bullet 0}^{+Y_2} \frac{\hat{p}_{10}^{(0)}}{\hat{p}_{\bullet 0}^{(0)}} = 20 + 15\frac{0{,}4}{0{,}6} + 15\frac{0{,}4}{0{,}5} = 42$$

$$E(n_{11} \mid n^{obs}, \theta^{(0)}) = n_{11}^R + n_{1\bullet}^{Y_1+} \frac{\hat{p}_{11}^{(0)}}{\hat{p}_{1\bullet}^{(0)}} + n_{\bullet 1}^{+Y_2} \frac{\hat{p}_{11}^{(0)}}{\hat{p}_{\bullet 1}^{(0)}} = 10 + 15\frac{0{,}2}{0{,}6} + 5\frac{0{,}2}{0{,}5} = 17\,.$$

Im anschließenden M-Schritt wird der Parametervektor $\theta^{(1)} = (\hat{p}_{00}^{(1)}, \hat{p}_{01}^{(1)}, \hat{p}_{10}^{(1)}, \hat{p}_{11}^{(1)})$ durch die (unter Vollständigkeit geltenden) ML-Schätzer der Multinomialverteilung

$$\hat{p}_{ab}^{(1)} = \frac{E(n_{ab} \mid n^{obs}, \theta^{(0)})}{n} = \frac{n_{ab}^R + n_{a\bullet}^{Y_1+} \frac{\hat{p}_{ab}^{(0)}}{\hat{p}_{a\bullet}^{(0)}} + n_{\bullet b}^{+Y_2} \frac{\hat{p}_{ab}^{(0)}}{\hat{p}_{\bullet b}^{(0)}}}{n} \qquad \forall\, a, b = 0, 1$$

neu bestimmt.[77] Dies führt im Beispiel zu den folgenden Schätzwerten:

$$\hat{p}_{00}^{(1)} = \frac{11{,}75}{100} = 0{,}1175 \quad \hat{p}_{01}^{(1)} = \frac{29{,}25}{100} = 0{,}2925 \quad \hat{p}_{10}^{(1)} = \frac{42}{100} = 0{,}42 \quad \hat{p}_{11}^{(1)} = \frac{17}{100} = 0{,}17$$

Die beiden Schritte des EM-Algorithmus werden solange wiederholt, bis eine Konvergenz bezüglich der Parameterschätzung eintritt. Im Beispiel ist die Bedingung

[76] Vgl. Schafer (1997), S. 44.
[77] Vgl. Schafer (1997), S. 44.

$$\left|\hat{p}_{ab}^{(t+1)} - \hat{p}_{ab}^{(t)}\right| \leq 0{,}0001 \qquad \forall\ a,b = 0{,}1$$

nach 10 Iterationen mit den Schätzern für die Zellwahrscheinlichkeiten

$$\hat{p}_{00}^{(1)} = 0{,}1343 \quad \hat{p}_{01}^{(1)} = 0{,}2839 \quad \hat{p}_{10}^{(1)} = \frac{42}{100} = 0{,}4229 \quad \hat{p}_{11}^{(1)} = \frac{17}{100} = 0{,}1589$$

erfüllt.

Nach Anwendung des EM-Algorithmus können die in der letzten Iteration generierten, vervollständigten Daten für die weitere statistische Analyse verwendet werden. Ein wesentlicher Nachteil dieser Vorgehensweise besteht darin, dass der Algorithmus lediglich die bedingten Erwartungswerte für die Ersetzung verwendet und somit die Unsicherheit bezüglich der Ergänzungen vernachlässigt wird. Somit sind zwar bei Gültigkeit der MAR-Annahme Parameterschätzungen basierend auf den vervollständigten Daten unverzerrt, die zugehörigen Standardfehler und Teststatistiken sind dagegen nicht verlässlich.[78, 79] Aus diesem Grund sind bayesianische Ansätze in Verbindung mit einer multiplen Ersetzung von fehlenden Werten, wie z.B. das Data-Augmentation Verfahren, vorzuziehen.

5.3.2 Data-Augmentation Verfahren

Wie bereits in Kapitel 4 erwähnt, werden innerhalb der bayesianischen Theorie alle Schlüsse für die unbekannten Parameter einer stetigen Verteilung aus der a posteriori Verteilung $P(\theta, \psi \mid \mathbf{y}_{obs}, \mathbf{r})$ gezogen. Für diese Verteilung gilt bei ignorierbarem Ausfallmechanismus:[80]

$$P(\theta \mid \mathbf{y}_{obs}, \mathbf{r}) = P(\theta \mid \mathbf{y}_{obs}) = \frac{P(\mathbf{y}_{obs} \mid \theta)\ P(\theta)}{\int P(\theta)\ P(\mathbf{y}_{obs} \mid \theta)\ d\theta}$$

Da die a posteriori Verteilung $P(\theta \mid \mathbf{y}_{obs})$ häufig nicht direkt bestimmbar ist, wird auf die leichter zu ermittelnde Verteilung $P(\theta \mid \mathbf{y}_{obs}, \mathbf{y}_{mis})$ zurückgegriffen. Hierfür werden, ähnlich wie beim EM-Algorithmus, in einem iterativen Verfahren die feh-

[78] Diese Problematik wurde bereits im Kapitel 5.2 im Rahmen der Multiplen Imputation aufgegriffen.
[79] Vgl. Statistical Services (2000).
[80] Vgl. Herleitung in Anhang A.3.

lenden Werte durch zufälliges Ziehen aus der Verteilung $P(\mathbf{Y}_{mis} \mid \mathbf{y}_{obs}, \theta^{(t)})$ unter Annahme eines Parameters $\theta^{(t)}$ in Iteration *t* provisorisch ersetzt:

$$\mathbf{y}_{mis}^{(t+1)} \sim P(\mathbf{Y}_{mis} \mid \mathbf{y}_{obs}, \theta^{(t)})$$

Im nächsten Schritt wird ein neuer Wert für θ aus der a posteriori Verteilung $P(\theta \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}^{(t+1)})$ gezogen:

$$\theta^{(t+1)} \sim P(\theta \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}^{(t+1)})$$

Es kann für hinreichend viele Iterationen *t* der beiden Schritte gezeigt werden, dass diese Vorgehensweise approximativ dem Ziehen aus der Verteilung $P(\theta, \mathbf{Y}_{mis} \mid \mathbf{y}_{obs})$ entspricht:[81]

$$\left(\theta^{(t)}, \mathbf{Y}_{mis}^{(t)}\right) \sim P(\theta, \mathbf{Y}_{mis} \mid \mathbf{y}_{obs})$$

Beispiel 5.5:[82]

Die eindimensionale Zufallsvariable Y_1 sei Bernoulli-verteilt mit Parameter θ:

$$Y_1 \sim B(1, \theta)$$

$$P(Y_{i1} = 0) = (1 - \theta)$$
$$P(Y_{i1} = 1) = \theta \qquad (i = 1, \ldots, n)$$

Die Wahrscheinlichkeit, die Realisationen ($y_1, \ldots, y_n$) unter dem Parameter θ zu beobachten, ist

$$P(y_{11}, \ldots, y_{1n} \mid \theta) = \theta^{\sum_{i=1}^{n} y_{i1}} (1 - \theta)^{n - \sum_{i=1}^{n} y_{i1}}.$$

Als a priori Verteilung von θ wird die Beta-Verteilung mit den Parametern α und β angenommen: [83]

---

[81] Vgl. Schafer (1997), S. 72.
[82] Vgl. Li (1988), S. 63; Schafer (1997), S. 76ff.
[83] Die Gleichverteilung ist ein Spezialfall der Beta-Verteilung mit $\alpha = 1$ und $\beta = 1$.

$$P(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \qquad \alpha>0, \beta>0 \tag{5.34}$$

Für die a posteriori Verteilung gilt:

$$P(\theta \mid y_{11},...,y_{n1}) = \frac{P(y_{11},...,y_{n1} \mid \theta)\ P(\theta)}{\int_0^1 P(y_{11},...,y_{n1} \mid \theta)\ P(\theta)\, d\theta}$$

$$= \frac{\theta^{\sum_{i=1}^{n} y_{i1}}(1-\theta)^{n-\sum_{i=1}^{n} y_{i1}} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)} \frac{\Gamma(\alpha+\sum_{i=1}^{n} y_{i1})\Gamma(\beta+n-\sum_{i=1}^{n} y_{i1})}{\Gamma(\alpha+\beta+n)}}$$

$$= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+\sum_{i=1}^{n} y_{i1})\Gamma(\beta+n-\sum_{i=1}^{n} y_{i1})}\theta^{\alpha-1+\sum_{i=1}^{n} y_{i1}}(1-\theta)^{\beta-1+n-\sum_{i=1}^{n} y_{i1}}$$

Die a posteriori Verteilung ist somit ebenfalls eine Beta-Verteilung mit den Parametern $\left(\alpha+\sum_{i=1}^{n} y_{i1}\right)$ und $\left(\beta+n-\sum_{i=1}^{n} y_{i1}\right)$:

$$\theta \mid y_{11},...,y_{n1} \sim Beta\left(\alpha+\sum_{i=1}^{n} y_{i1}, \beta+n-\sum_{i=1}^{n} y_{i1}\right)$$

Es wird angenommen, dass die ersten *q* Merkmalsträger beobachtet wurden, während die Werte der restlichen (*n-q*) Merkmalsträger nicht bekannt sind. Für die Parameter der a priori Verteilung in (5.34) soll $\alpha \to 0$ und $\beta \to 0$ angenommen werden.[84] Dann gilt für die a posteriori Verteilung von θ gegeben die beobachteten Daten $\mathbf{y}_{obs} = (y_{11},...,y_{q1})$:

$$\theta \mid \mathbf{y}_{obs} \sim Beta\left(\sum_{i=1}^{q} y_{i1}, n-\sum_{i=1}^{q} y_{i1}\right)$$

[84] Vgl. Li (1988), S. 63.

Im ersten Schritt des Data-Augmentation-Verfahrens werden die fehlenden Werte $\mathbf{y}_{mis} = (y_{(q+1)1},...,y_{n1})$ aus der folgenden Verteilung mit dem beliebig festzulegenden Startwert $\theta^{(0)}$ gezogen:

$$P(Y_{i1}^{(1)} = 0) = (1 - \theta^{(0)})$$
$$P(Y_{i1}^{(1)} = 1) = \theta^{(0)} \qquad (i = 1,...,n)$$

Im nächsten Schritt wird der Parameter $\theta^{(1)}$ aus der a posteriori Verteilung $P(\theta \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}^{(1)})$ bestimmt:

$$\theta^{(1)} \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}^{(1)} \sim Beta\left(\sum_{i=1}^{q} y_{i1} + \sum_{i=q+1}^{n} y_{i1}^{(1)}, n - \sum_{i=1}^{q} y_{i1} - \sum_{i=q+1}^{n} y_{i1}^{(1)} \right)$$

Für den bedingten Erwartungswert eines fehlenden Wertes in Iteration *t*+1 gegeben die beobachteten und in *t* ersetzten Daten folgt:[85]

$$\begin{aligned} E(Y_1^{(t+1)} \mid \mathbf{y}_{mis}^{(t)}, \mathbf{y}_{obs}) &= E(\theta^{(t)} \mid \mathbf{y}_{mis}^{(t)}, \mathbf{y}_{obs}) \\ &= \frac{\sum_{i=1}^{q} y_{i1} + \sum_{i=q+1}^{n} y_{i1}^{(t)}}{n} \\ &= \frac{\sum_{i=1}^{q} y_{i1}}{q} - \varsigma \left(\frac{\sum_{i=q+1}^{n} y_{i1}^{(t)}}{n-q} - \frac{\sum_{i=1}^{q} y_{i1}}{q} \right) \qquad ; \quad \varsigma = \frac{n-q}{n} \end{aligned}$$

Der bedingte Erwartungswert eines fehlenden Wertes in Iteration *t*+1 gegeben die beobachteten und der in der *ersten* Iteration ersetzten Daten ist

$$E(Y_1^{(t+1)} \mid \mathbf{y}_{mis}^{(1)}, \mathbf{y}_{obs}) = \frac{\sum_{i=1}^{q} y_{i1}}{q} - (\varsigma)^t \left(\frac{\sum_{i=q+1}^{n} y_{i1}^{(1)}}{n-q} - \frac{\sum_{i=1}^{q} y_{i1}}{q} \right) \quad .$$

[85] Vgl. Schafer (1997), S. 77.

Da die in Iteration 1 ersetzten Daten nur vom Parameter $\theta^{(0)}$ abhängen, gilt:[86]

$$E(Y_1^{(t+1)} \mid \theta^{(0)}, \mathbf{y}_{obs}) = \frac{\sum_{i=1}^{q} y_{i1}}{q} - (\varsigma)^t \left(\gamma - \frac{\sum_{i=1}^{q} y_{i1}}{q} \right) \tag{5.35}$$

Für jeden Startwert γ geht der Erwartungswert in (5.35) für $t \to \infty$ gegen $\frac{\sum_{i=1}^{q} y_{i1}}{q}$, da $\zeta < 1$ ist. Dies entspricht dem bedingten Erwartungswert eines fehlenden Wertes in Iteration t+1 gegeben die beobachteten Daten $\mathbf{y}_{obs}$:

$$E(Y_1^{(t+1)} \mid \mathbf{y}_{obs}) = E(\theta^{(t)} \mid \mathbf{y}_{obs}) = \frac{\sum_{i=1}^{q} y_{i1}}{q}$$

Somit ist in dem betrachteten Beispiel sowohl $E(Y_1^{(t+1)} \mid \mathbf{y}_{obs})$ als auch der dem Data-Augmentation-Verfahren zugrunde liegende, bedingte Erwartungswert $E(Y_1^{(t+1)} \mid \theta^{(0)}, \mathbf{y}_{obs})$ das arithmetische Mittel der beobachteten Daten. Die in dem Algorithmus iterativ durchgeführte (provisorische) Ersetzung der fehlenden Werte und die daran anschließende Ermittlung von $\theta^{(t)}$ führt zum wahren Erwartungswert eines fehlenden Wertes.

Eine Erweiterung des Data-Augmentation-Verfahrens besteht in dem m-maligen Ziehen ($m > 1$) von Werten aus der prädiktiven bedingten Verteilung von $\mathbf{Y}_{mis}$ in jeder Iteration. Diese multiple Imputation führt zu m verschiedenen, vervollständigten Datenbeständen in Iteration (t+1):

$$\mathbf{y}_{mis,h}^{(t+1)} \sim P(\mathbf{Y}_{mis} \mid \mathbf{y}_{obs}, \theta_h^{(t)}) \qquad (1 \le h \le m)$$

[86] Vgl. Schafer (1997), S. 77.

Dabei ist $\theta_h^{(t)}$ der jeweilige Parameterschätzwert für Datenbestand *h* des vorherigen Durchlaufs, der im nächsten Schritt durch zufälliges Ziehen aus der a posteriori Verteilung neu ermittelt wird:

$$\theta_h^{(t+1)} \sim P(\theta \mid \mathbf{y}_{obs}, \mathbf{y}_{mis,h}^{(t+1)}) \qquad (1 \le h \le m)$$

Jeder der *m* vervollständigten Datenbestände wird anschließend mit den entsprechenden statistischen Methoden, die auch bei vollständiger Beobachtung aller Merkmalsträger anzuwenden wären, analysiert. Die einzelnen Ergebnisse der Analysen werden abschließend durch die Bildung des arithmetischen Mittels zu einem endgültigen Schätzer zusammengefasst.

Durch die mehrfache Ersetzung der fehlenden Werte wird die Unsicherheit bezüglich des unbekannten, wahren Ausfallmechanismus innerhalb der Behandlungsmethode berücksichtigt.[87] Die multiple Imputation von fehlenden Werten ermöglicht es, die Varianz der geschätzten Statistiken exakt zu bestimmen und ist somit der einfachen Ersetzung vorzuziehen.

[87] Vgl. Little/Rubin (2002), S. 85.

6 Behandlungsverfahren bei nicht ignorierbaren Ausfallmechanismen

6.1 Problemdarstellung

Die Ausführungen im letzten Kapitel beschränkten sich auf Fälle, bei denen Daten zufällig (MCAR) bzw. systematisch im Sinne von MAR fehlten. Unter dieser Voraussetzung[88] ist es möglich, den Ausfallmechanismus für die Schätzung von Parametern aus den beobachteten Daten zu vernachlässigen. Die in Kapitel 5 vorgestellten Methoden zur Behandlung von Datenausfall beruhen dabei auf der prinzipiellen Annahme, dass die systematischen Unterschiede zwischen beobachteten und fehlenden Werten einer Variable allein durch die Kovariaten erklärt werden können.[89]

In der Praxis kann die Gültigkeit der MAR-Annahme jedoch nicht immer vorausgesetzt werden, wie z.B. empirische Studien von Greenlees et al. (1982) und Smith et al. (1999) belegen. Wie bereits in Kapitel 3 diskutiert, vertreten andererseits zahlreiche Autoren die Auffassung, dass die MAR-Annahme bei umfangreichen, multivariaten Datenbeständen fast immer erfüllt ist. Selbst für den Fall, dass diese strittige und allgemein nicht überprüfbare Aussage Bestand hat, ist festzuhalten, dass zumindest bei der Analyse von Datenbeständen mit einer geringen Anzahl an Variablen die Notwendigkeit besteht, nicht ignorierbare Ausfallmechanismen bei der Behandlung in Betracht zu ziehen.

Die Ergänzung fehlender Werte bei Vorliegen von MNAR erweist sich als kompliziert, da die Schlüsse bezüglich des Parametervektors θ – im Gegensatz zu MAR – nicht aus der Wahrscheinlichkeitsfunktion $P_\theta(\mathbf{y}_{obs})$ der beobachteten Daten gezogen werden können. Stattdessen ist – aufgrund der Nichtignorierbarkeit des Ausfallmechanismus – der Parametervektor θ aus der gemeinsamen Wahrscheinlichkeitsfunktion $P_{\theta,\psi}(\mathbf{y}_{obs}, \mathbf{r})$ zu schätzen, indem die entsprechende Likelihood-Funktion $L(\mathbf{y}_{obs}, \mathbf{r} \mid \theta, \psi)$ maximiert wird. Prinzipiell ist bei diesen Verfahren die Spezifizierung des zugrunde liegenden Ausfallmechanismus erforderlich. Dieser hängt auch

[88] Die Unabhängigkeit der Parametervektoren θ und ψ sei im Folgenden vorausgesetzt.
[89] Vgl. Allison (2002), S. 77.

von den nicht beobachteten Werten von **y** ab, so dass die Parameter der Verteilung von **R** gegeben **y** bzw. der Verteilungstyp selbst nicht ohne zusätzliche Annahmen bestimmt werden können. Damit verbunden ist die Problematik, dass – ohne zusätzliches Wissen – *alle* denkbaren Ausfallmodelle (sowohl ignorierbarer als auch nicht ignorierbarer Art) plausibel sind und es kein Kriterium für die Präferenz eines Modells gegenüber den Alternativen gibt.

Da die Fehlspezifikation eines Modells für den Datenausfall erhebliche Auswirkungen auf die Schätzung von Parametern hat, ist die Sensitivität bezüglich des Ergebnisses der statistischen Analyse durch Modellierung mehrerer plausibler Ausfallmechanismen zu untersuchen. Das Ziel ist somit die Angabe eines Intervalls für den unbekannten Parameter statt einer Punktschätzung.[90]

Im Folgenden werden zunächst zwei spezielle Verfahren vorgestellt, die lediglich bei kategorialen Merkmalen anwendbar sind. Allgemeine Ansätze zur Behandlung von fehlenden Werten werden unter der Prämisse, dass der Ausfallmechanismus vom Typ MNAR ist, in den Kapiteln 6.3 und 6.4 diskutiert.

6.2 Spezielle Verfahren für diskrete Variablen

6.2.1 Zielsetzung

In diesem Abschnitt werden zwei Verfahren betrachtet, die zur Behandlung von fehlenden Werten angewendet werden können, falls die betrachteten Zufallsvariablen diskret sind und der Ausfallmechanismus nicht ignorierbar ist. In Kapitel 6.2.2 wird eine Methode beschrieben, deren eigentlicher Anwendungsbereich in der Klassifikation von Probanden liegt, wobei diese Problemstellung aus dem medizinischen Bereich auf die Modellierung eines Ausfallmechanismus vom Typ MNAR zurückgeführt und somit in allgemeiner Form dargestellt werden kann. Innerhalb des Verfahrens wird eine Zufallsvariable Y_1 betrachtet, deren (Nicht-)Beobachtung in einer Stichprobe auf deren Realisation y_1 zurückzuführen ist:

$$P(R_1 = r_1 \mid Y_1 = y_1) \neq P(R_1 = r_1)$$

[90] Vgl. Allison (2002), S. 78.

Die Anwendung der Methode ermöglicht es, unter diesem MNAR-Ausfallmechanismus sowie weiteren zu treffenden Annahmen die Parameter der marginalen Verteilung $P(Y_1)$ zu schätzen. Unterschiedliche Annahmen führen dabei zu verschiedenen Parameterschätzungen, so dass mit Hilfe des Verfahrens eine Sensitivitätsbetrachtung bezüglich des Schätzers durchgeführt werden kann.

Den Ausgangspunkt des in Kapitel 6.2.3 diskutierten Verfahrens stellen zwei diskrete Zufallsvariablen Y_1 und Y_2 dar, wobei annahmegemäß lediglich Y_2 fehlende Werte in einer Stichprobe aufweist. Es wird gezeigt, dass bei Vorliegen eines nicht ignorierbaren Ausfallmechanismus

$$P(R_2 = r_2 \mid Y_1 = y_1, Y_2 = y_2) \neq P(R_2 = r_2 \mid Y_1 = y_1)$$

eine unverzerrte Schätzung der Parameter der gemeinsamen Verteilung $P(Y_1, Y_2)$ möglich ist, falls die bedingte Unabhängigkeitsbeziehung

$$R_2 \perp\!\!\!\perp Y_1 \mid Y_2 \tag{6.1}$$

gilt. Das Verfahren wird in einer Simulationsstudie (Kapitel 7.2) zur Parameterschätzung angewendet, in der u.a. für den Fall, dass die bedingte Unabhängigkeitsbeziehung (6.1) verletzt ist, die Robustheit der Methode untersucht wird.

6.2.2 Sensitivitätsanalyse der Parameterschätzung bei einer diskreten Variable

Eine spezielle Form des Datenausfalls liegt vor, falls anhand von (vollständig erhobenen) Daten Merkmalsträger klassifiziert werden sollen und diese Einteilung nicht in allen Fällen eindeutig möglich ist. Neben den für die Klassifikation notwendigen Ausprägungen ist somit eine weitere Kategorie für nicht zuordenbare Merkmalsträger zu berücksichtigen. Hängt die Wahrscheinlichkeit für die Nichtzuordnung nur von der tatsächlichen Kategorie ab, so kann die Problemstellung auf einen Datenausfall vom Typ MNAR zurückgeführt werden. Typische Beispiele für diese Klassifikationsproblematik sind im medizinischen Bereich zu finden, bei denen anhand von Symptomen eine Einteilung vorzunehmen ist. Die folgenden Ausführungen sind insbesondere auf Nordheim (1984) zurückzuführen, der bei dieser Form des Datenausfalls eine Behandlung von fehlenden Werten innerhalb einer medizinischen Studie zeigte.

Das Vorhandensein eines Symptoms kann durch eine Zufallsvariable Y_1 beschrieben werden. Anhand einer Untersuchung wird der Status „Symptom vorhanden“ bzw. „Symptom nicht vorhanden“ bei den Probanden festgestellt, wobei einige Personen nicht klassifiziert werden konnten. Es wird zunächst angenommen, dass die Probanden korrekt zugewiesen werden und somit keine Fehlklassifikation existiert. Die eindimensionale Zufallsvariable Y_1 ist eine binäre Zufallsvariable, deren Realisationen in einer Stichprobe vom Umfang n nicht vollständig beobachtet wurden. Weiterhin ist n_a^R die Anzahl der beobachteten Realisationen mit $Y_1 = a$ (a = 0, 1) und n^{NR} die Anzahl der fehlenden Werte in der Stichprobe. Die Wahrscheinlichkeiten $P_{\theta_a}(Y_1 = a) = \theta_a$ (a = 0, 1) sind aufgrund der fehlenden Werte unbekannt und sollen geschätzt werden ($\theta = (\theta_0, \theta_1)$, $\theta_1 = 1\text{-}\theta_0$). Aus dem gleichen Grund sind die bedingten Wahrscheinlichkeiten $P_{\psi_a}(R_1 = 0 \mid Y_1 = a) = \psi_a$ (a = 0, 1) ebenfalls nicht bekannt ($\psi = (\psi_0, \psi_1)$).

Die beobachteten Werte der binären Zufallsvariable Y_1 werden durch die Häufigkeiten n_0^R, n_1^R und n^{NR} vollständig beschrieben. Diese Häufigkeiten sind Realisationen der mehrdimensionalen Zufallsvariable (N_0^R, N_1^R, N^{NR}), die multinomialverteilt ist mit den Parametern n, $P_{\theta_0,\psi_0}(R_1 = 0, Y_1 = 0)$, $P_{\theta_1,\psi_1}(R_1 = 0, Y_1 = 1)$ und $P_{\theta,\psi}(R_1 = 1)$:[91]

$$P_{\theta,\psi}(N_0^R = n_0^R, N_1^R = n_1^R, N^{NR} = n^{NR}) = \frac{n!}{n_0^R!\, n_1^R!\, n^{NR}!} P_{\theta_0,\psi_0}(R = 0, Y = 0)^{n_0^R} \cdot P_{\theta_1,\psi_1}(R = 0, Y = 1)^{n_1^R} P_{\theta,\psi}(R = 1)^{n^{NR}} \qquad (6.2)$$

Durch Umformen von (6.2) kann die Wahrscheinlichkeit

$$P_{\theta,\psi}(N_0^R = n_0^R, N_1^R = n_1^R, N^{NR} = n^{NR})$$

in Abhängigkeit von den Parametern θ_0, ψ_0 und ψ_1 angegeben werden:

[91] Vgl. Nordheim (1984), S. 774.

$$P_{\theta,\psi}(N_0^R = n_0^R, N_1^R = n_1^R, N^{NR} = n^{NR}) = \frac{n!}{n_0^R!\,n_1^R!\,n^{NR}!}\Big(P_{\theta_0}(Y_1=0)P_{\psi_0}(R_1=0\mid Y_1=0)\Big)^{n_0^R}$$
$$\cdot\Big(P_{\theta_1}(Y_1=1)P_{\psi_1}(R_1=0\mid Y_1=1)\Big)^{n_1^R}\Big(P_{\theta_0,\psi_0}(Y_1=0,R_1=1)+P_{\theta_1,\psi_1}(Y_1=1,R_1=1)\Big)^{n^{NR}}$$

$$=\frac{n!}{n_0^R!\,n_1^R!\,n^{NR}!}\Big(P_{\theta_0}(Y_1=0)P_{\psi_0}(R_1=0\mid Y_1=0)\Big)^{n_0^R}\Big((1-P_{\theta_0}(Y_1=0))P_{\psi_1}(R_1=0\mid Y_1=1)\Big)^{n_1^R}$$
$$\cdot\Big(P_{\theta_0}(Y_1=0)(1-P_{\psi_0}(R_1=0\mid Y_1=0))+(1-P_{\theta_0}(Y_1=0))(1-P_{\psi_1}(R_1=0\mid Y_1=1))\Big)^{n^{NR}}$$

$$=\frac{n!}{n_0^R!\,n_1^R!\,n^{NR}!}(\theta_0\psi_0)^{n_0^R}\big((1-\theta_0)\psi_1\big)^{n_1^R}\big(\theta_0(1-\psi_0)+(1-\theta_0)(1-\psi_1)\big)^{n^{NR}} \tag{6.3}$$

Die Schätzung des Parameters θ_0 aus (6.3) ist bei unbekanntem Ausfallmechanismus nur möglich, falls eine Restriktion bezüglich der Parameter ψ_0 und ψ_1 hinzugefügt wird. Diese Bedingung kann anhand einer Ratio K formuliert werden, welche wie folgt definiert ist:

$$K=\frac{P_{\psi_1}(R=1\mid Y=1)}{P_{\psi_0}(R=1\mid Y=0)}=\frac{1-\psi_1}{1-\psi_0} \tag{6.4}$$

Falls $K > 1$ gilt, so ist die Wahrscheinlichkeit, dass Y_1 nicht beobachtet wird, wenn $Y_1 = 1$ ist, größer als im Fall $Y_1 = 0$. Unter Berücksichtigung von (6.4) gilt für die gemeinsame Wahrscheinlichkeit von (N_0^R, N_1^R, N^{NR}) in (6.3):[92]

$$P(N_0^R = n_0^R, N_1^R = n_1^R, N^{NR} = n^{NR}) = \frac{n!}{n_0^R!\,n_1^R!\,n^{NR}!}L_I(\theta_0,K)\,L_{II}(\psi_0,K)$$

$$L_I(\theta_0,K)=(\theta_0)^{n_0^R}(1-\theta_0)^{n_1^R}\big(\theta_0+K(1-\theta_0)\big)^{n^{NR}}$$

$$L_{II}(\psi_0,K)=(\psi_0)^{n_0^R}\big((1-K(1-\psi_0)\big)^{n_1^R}(1-\psi_0)^{n^{NR}} \tag{6.5}$$

Wird die Ratio K als bekannt angenommen, so sind Rückschlüsse bezüglich des zu schätzenden Parameters θ_0 allein aus $L_I(\theta_0,K)$ zu ziehen, da der Term $L_{II}(\psi_0,K)$ nicht von θ_0 abhängig ist. Demnach gilt für die Loglikelihood-Funktion zu der Multinomialverteilung in (6.5) gegeben θ_0:

$$l\big(n_0^R, n_1^R, n^{NR}\mid\theta_0\big)\propto n_0^R\ln(\theta_0)+n_1^R\ln(1-\theta_0)+n^{NR}\ln(\theta_0+K(1-\theta_0))$$

[92] Vgl. Nordheim (1984), S. 774.

Die Maximierung dieser Loglikelihood-Funktion nach θ_0 führt zu der folgenden Gleichung:

$$\left(\hat{\theta}_0\right)^2 n(1-K) + \hat{\theta}_0\left(K(n+n_0^R) - (n_0^R + n^{NR})\right) - Kn_0^R = 0 \tag{6.6}$$

Die beiden Lösungen der quadratischen Gleichung ergeben sich durch die folgende Formel:

$$\hat{\theta}_{0(1,2)} = -\frac{K(n+n_0^R) - (n_0^R + n^{NR})}{2n(1-K)} \pm \sqrt{\left(\frac{K(n+n_0^R) - (n_0^R + n^{NR})}{\frac{n(1-K)}{2}}\right)^2 + \frac{Kn_0^R}{n(1-K)}} \tag{6.7}$$

Die quadratische Gleichung ist für den Fall lösbar, dass die absoluten Zellhäufigkeiten von der Zufallsvariable Y_1 positiv sind $(n_0^R, n_1^R, n^{NR} > 0)$ und $K \neq 1$ ist. Eine der beiden Lösungen liegt außerhalb des zulässigen Parameterbereichs [0,1] von θ_0, während die andere Lösung durch die relativen Häufigkeiten $\frac{n_0^R}{n}$ und $\frac{n_0^R + n^{NR}}{n}$ begrenzt wird und damit der ML-Schätzer von θ_0 ist:

$$\frac{n_0^R}{n} \leq \hat{\theta}_0 \leq \frac{n_0^R + n^{NR}}{n} \tag{6.8}$$

Beispiel 6.1:

In einer Stichprobe vom Umfang n = 1000 seien 200 mal der Wert 0 $(n_0^R = 200)$ und 300 mal der Wert 1 $(n_1^R = 300)$ einer eindimensionalen, binären Zufallsvariable Y_1 beobachtet worden. In 500 Fällen konnte die Realisation von Y_1 nicht bestimmt werden $(n^{NR} = 500)$. Ausgehend von Formel (6.7) kann der ML-Schätzer von θ_0 unter Annahme verschiedener Werte von K ermittelt werden. Der Zusammenhang zwischen den Schätzwerten $\hat{\theta}_0$ und K ist in der folgenden Abbildung graphisch dargestellt.

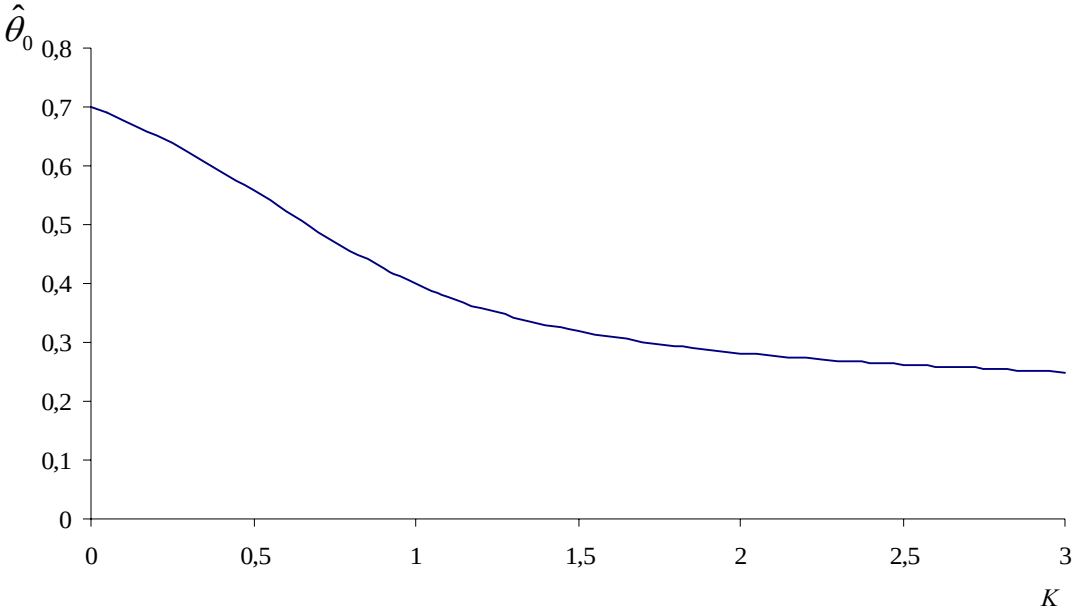


Abbildung 6.1: ML-Schätzer $\hat{\theta}_0$ in Abhängigkeit von K (Beispiel 6.1)

Der ML-Schätzer von θ_0 ist im Extremfall 0,7, da unter $K = 0$ für die bedingte Ausfallwahrscheinlichkeit $P_{\psi_1}(R_1 = 1 \mid Y_1 = 1) = 0$ gilt und allen fehlenden Werten der Wert Null zugewiesen wird. Demzufolge wird für den Schätzwert $\hat{\theta}_0$ die in (6.8) angegebene Obergrenze erreicht:

$$\hat{\theta}_0 = \frac{n_0^R + n^{NR}}{n} = \frac{200 + 500}{1000} = 0{,}7$$

Für $K \to \infty$ werden die nicht beobachteten Daten durch den Wert Eins ersetzt, so dass für $\hat{\theta}_0$ gilt:

$$\hat{\theta}_0 = \frac{n_0^R}{n} = \frac{200}{1000} = 0{,}2 \tag{6.9}$$

Beispiel 6.2:

Abweichend vom vorigen Beispiel seien die folgenden Häufigkeiten der Zufallsvariable Y_1 erhoben worden:

$$n_0^R = 100 \quad , \quad n_1^R = 150 \quad , \quad n^{NR} = 750$$

Die relativen Häufigkeiten der beobachteten Daten

$$\frac{n_0^R}{n}, \frac{n_1^R}{n}$$

sind gegenüber dem vorigen Szenario gleich geblieben, die Ausfallquote

$$\frac{n^{NR}}{n} \cdot 100\%$$

ist jedoch mit 75% wesentlich höher als in Beispiel 6.1 mit 50%. Dies hat die folgenden Auswirkungen auf den ML-Schätzer $\hat{\theta}_0$:

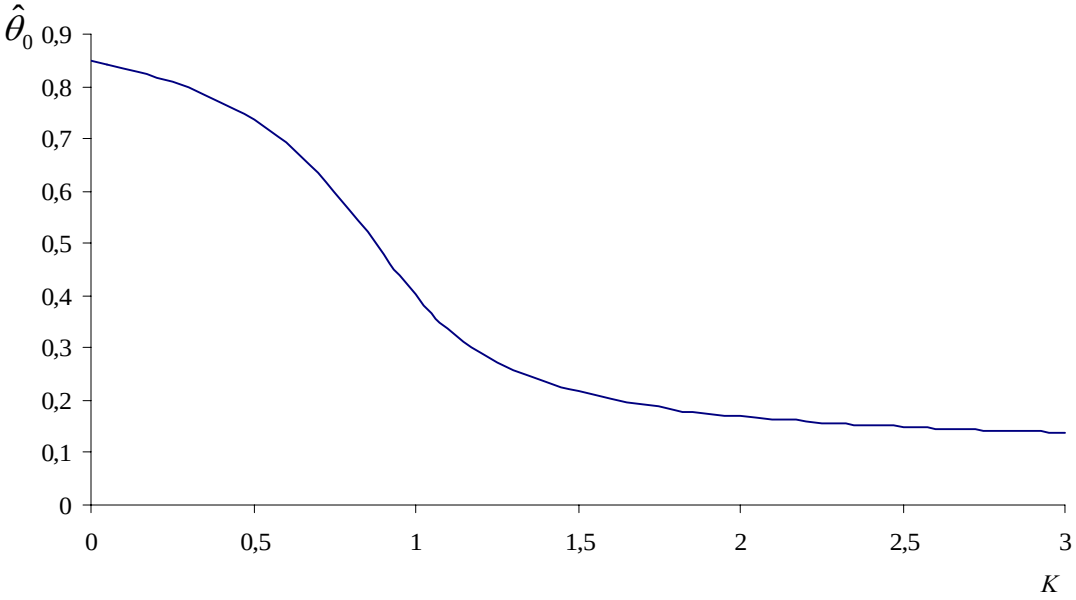


Abbildung 6.2: ML-Schätzer $\hat{\theta}_0$ in Abhängigkeit von K (Beispiel 6.2)

Der Parameter θ_0 liegt im Intervall [0,1; 0,85], das aufgrund der größeren Anzahl fehlender Werte breiter ist als im vorherigen Beispiel. Eine größere Anzahl von Be-

obachtungen in der Stichprobe führt somit ceteris paribus zu einem kleineren Intervall und ermöglicht genauere Aussagen bezüglich des Parameters θ_0.

Im Fall von $K = 1$ sind die bedingten Ausfallwahrscheinlichkeiten in (6.4) gleich:

$$P_{\psi_0}(R_1 = 1 \mid Y_1 = 0) = P_{\psi_1}(R_1 = 1 \mid Y_1 = 1)$$

Der Datenausfall ist nicht von Y abhängig, so dass die MCAR-Annahme und auch die MAR-Annahme erfüllt sind. Letzteres gilt, da in diesem Abschnitt keine Kovariaten betrachtet werden. Unter $K = 1$ vereinfacht sich (6.6) zu der folgenden Gleichung:

$$\hat{\theta}_0\left((n + n_0^R) - (n_0^R + n^{NR})\right) - n_0^R = 0$$
$$\Leftrightarrow \hat{\theta}_0 = \frac{n_0^R}{n - n^{NR}} = \frac{n_0^R}{n_0^R + n_1^R}$$

Unter Gültigkeit der MAR-Annahme ist der Schätzwert $\hat{\theta}_0$ somit der relative Anteil der beobachteten Werte mit $Y_1 = 0$ unter allen Beobachtungen. Aus dem Beispiel 6.1 ist ersichtlich, dass unter diesem Szenario Abweichungen von der MAR-Annahme ($K = 1$) geringere Auswirkungen auf die Parameterschätzungen haben als im Beispiel 6.2. Dies ist wiederum auf den hohen Anteil fehlender Werte in letzterem Beispiel zurückzuführen. Insbesondere in diesen Fällen ist es daher notwendig, den MNAR-Ausfall in Betracht zu ziehen und unter Berücksichtigung verschiedener Mechanismen eine Sensitivitätsanalyse durchzuführen. Um den Wertebereich von $\hat{\theta}_0$ weiter einschränken zu können, ist die Einbeziehung von Expertenwissen bezüglich der Ratio K hilfreich. So kann z.B. in den meisten medizinischen Studien davon ausgegangen werden, dass es wahrscheinlicher ist, eine Person klassifizieren zu können, wenn das Symptom vorhanden ist ($Y_1 = 1$), als eine Person klassifizieren zu können, wenn ein Symptom nicht vorhanden ist ($Y_1 = 0$):[93]

[93] Vgl. Nordheim (1984), S. 772f.

$$P_{\psi_1}(R_1 = 0 \mid Y_1 = 1) > P_{\psi_0}(R_1 = 0 \mid Y_1 = 0)$$
$$\Leftrightarrow P_{\psi_1}(R_1 = 1 \mid Y_1 = 1) < P_{\psi_0}(R_1 = 1 \mid Y_1 = 0) \tag{6.10}$$

Unter den Personen, die nicht klassifiziert werden konnten, befindet sich somit ein größerer Anteil Probanden, die das Symptom nicht aufweisen. In diesen Fällen kann ein kleineres Intervall für $\hat{\theta}_0$ als in (6.8) angegeben werden, da aufgrund von Ungleichung (6.10) für K gilt:

$$0 \leq K < 1 \tag{6.11}$$

Im Beispiel 6.1 ergibt sich unter der Voraussetzung (6.10) für den Schätzwert $\hat{\theta}_0$ das Intervall]0,4; 0,7] mit einer höheren Untergrenze als in (6.9). Häufig kann sogar dieser Bereich noch weiter eingeschränkt werden, da unter Einbeziehung von Experteninformationen der Wert für K noch stärker als in (6.11) begrenzt werden kann.

Nordheim (1984) erweiterte das beschriebene Modell, so dass mehr als zwei Kategorien von Y_1 möglich sind bzw. Fehlklassifikationen auftreten können. Diese Erweiterungen setzen jedoch voraus, dass – neben der Ratio K – weitere Größen definiert und geschätzt werden müssen. Die Maximierung der Likelihood-Funktion bezüglich des Parameters θ_0 ist in diesem erweiterten Modell analog zu der beschriebenen Methodik durchzuführen, und die Sensitivität der Schätzung kann anhand der Variierung der vorzugebenden Größen ermittelt werden. Aufgrund der weitgehend identischen Vorgehensweisen sei an dieser Stelle auf Nordheim (1984) für die weiteren Ausführungen verwiesen.[94]

6.2.3 Parameterschätzungen bei zwei binären Variablen

Im vorangegangenen Kapitel erfolgte die Parameterschätzung für den Fall, dass lediglich eine Variable betrachtet wird, deren Ausfall durch einen nicht ignorierbaren Mechanismus verursacht worden ist. Ausgangspunkt des folgenden Verfahrens sind n Untersuchungseinheiten von zwei binären Variablen Y_1 und Y_2, wobei in der Stich-

[94] Vgl. Nordheim (1984), S. 776 f.

probe Y_1 vollständig beobachtet wurde und Y_2 fehlende Werte aufweist. Weiterhin wird vorausgesetzt, dass R_2 und Y_1 gegeben Y_2 unabhängig sind:

$$R_2 \perp\!\!\!\perp Y_1 \mid Y_2 \tag{6.12}$$

Für die Verteilung $P_\psi(R_2 \mid y_1, y_2)$ gilt in diesem Fall ($\psi = (\psi_1, \psi_2)$):

$$P_{\psi_1}(R_2 = 0 \mid Y_1 = y_1, Y_2 = 0) = P_{\psi_1}(R_2 = 0 \mid Y_2 = 0) = \psi_1$$
$$P_{\psi_2}(R_2 = 0 \mid Y_1 = y_1, Y_2 = 1) = P_{\psi_2}(R_2 = 0 \mid Y_2 = 1) = \psi_2$$

Die Verteilung $P_\theta(Y_1, Y_2)$ wird durch den Parametervektor $\theta = (\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11})$ in der folgenden Weise beschrieben:

$$P_{\theta_{ab}}(Y_1 = a, Y_2 = b) = \theta_{ab} \qquad \forall a, b = 0,1$$

Baker/Laird (1988) zeigen unter Anwendung eines loglinearen Modells, dass der Parametervektor θ der Verteilung $P_\theta(Y_1, Y_2)$ bei diesem Ausfallmechanismus vom Typ MNAR unverzerrt geschätzt werden kann.[95] Eine alternative Herleitung dieser Schätzer ist über den Zusammenhang

$$P(Y_1 = 0 \mid Y_2 = 0, R_2 = 0) = P(Y_1 = 0 \mid Y_2 = 0, R_2 = 1)$$

$$\Rightarrow \frac{P(Y_1 = 0, Y_2 = 0, R_2 = 0)}{P(Y_2 = 0, R_2 = 0)} = \frac{P(Y_1 = 0, Y_2 = 0, R_2 = 1)}{P(Y_2 = 0, R_2 = 1)} \tag{6.13}$$

möglich, welcher unmittelbar aus der Annahme in (6.12) folgt.

Analog gilt für die bedingten Wahrscheinlichkeiten $P(Y_1 = 1 \mid Y_2 = 0, R_2 = 0)$ und $P(Y_1 = 1 \mid Y_2 = 0, R_2 = 1)$:

$$\frac{P(Y_1 = 1, Y_2 = 0, R_2 = 0)}{P(Y_2 = 0, R_2 = 0)} = \frac{P(Y_1 = 1, Y_2 = 0, R_2 = 1)}{P(Y_2 = 0, R_2 = 1)} \tag{6.14}$$

Aus (6.13) und (6.14) folgt:

$$\frac{P(Y_1 = 0, Y_2 = 0, R_2 = 0)}{P(Y_1 = 0, Y_2 = 0, R_2 = 1)} = \frac{P(Y_1 = 1, Y_2 = 0, R_2 = 0)}{P(Y_1 = 1, Y_2 = 0, R_2 = 1)} \tag{6.15}$$

In gleicher Weise kann gezeigt werden, dass

---

[95] Vgl. Baker/Laird (1988), S. 67.

$$\frac{P(Y_1 = 0, Y_2 = 1, R_2 = 0)}{P(Y_1 = 0, Y_2 = 1, R_2 = 1)} = \frac{P(Y_1 = 1, Y_2 = 1, R_2 = 0)}{P(Y_1 = 1, Y_2 = 1, R_2 = 1)} \tag{6.16}$$

gilt.

Aus dem Gleichungssystem bestehend aus den Formeln (6.15), (6.16) sowie den Wahrscheinlichkeiten

$$P(Y_1 = 0, R_2 = 1) = P(Y_1 = 0, Y_2 = 0, R_2 = 1) + P(Y_1 = 0, Y_2 = 1, R_2 = 1) \tag{6.17}$$

$$P(Y_1 = 1, R_2 = 1) = P(Y_1 = 1, Y_2 = 0, R_2 = 1) + P(Y_1 = 1, Y_2 = 1, R_2 = 1) \tag{6.18}$$

kann die gemeinsame Wahrscheinlichkeit $P(Y_1 = a, Y_2 = b, R_2 = 1)$ $(\forall a, b = 0,1)$ ermittelt werden:

$$P(Y_1 = 0, Y_2 = 0, R_2 = 1) = \frac{P(Y_1 = 1, R_2 = 1) - \dfrac{P(Y_1 = 0, R_2 = 1)\, P(Y_1 = 1, Y_2 = 1, R_2 = 0)}{P(Y_1 = 0, Y_2 = 1, R_2 = 0)}}{\dfrac{P(Y_1 = 1, Y_2 = 0, R_2 = 0)}{P(Y_1 = 0, Y_2 = 0, R_2 = 0)} - \dfrac{P(Y_1 = 1, Y_2 = 1, R_2 = 0)}{P(Y_1 = 0, Y_2 = 1, R_2 = 0)}} \tag{6.19}$$

$$P(Y_1 = 1, Y_2 = 0, R_2 = 1) = \frac{P(Y_1 = 1, R_2 = 1) - \dfrac{P(Y_1 = 0, R_2 = 1)\, P(Y_1 = 1, Y_2 = 1, R_2 = 0)}{P(Y_1 = 0, Y_2 = 1, R_2 = 0)}}{1 - \dfrac{P(Y_1 = 1, Y_2 = 1, R_2 = 0) P(Y_1 = 0, Y_2 = 0, R_2 = 0)}{P(Y_1 = 0, Y_2 = 1, R_2 = 0) P(Y_1 = 1, Y_2 = 0, R_2 = 0)}} \tag{6.20}$$

$$P(Y_1 = 0, Y_2 = 1, R_2 = 1) = P(Y_1 = 0, R_2 = 1) - P(Y_1 = 0, Y_2 = 0, R_2 = 1) \tag{6.21}$$

$$P(Y_1 = 1, Y_2 = 1, R_2 = 1) = P(Y_1 = 1, R_2 = 1) - P(Y_1 = 1, Y_2 = 0, R_2 = 1) \tag{6.22}$$

Diese Gleichungen implizieren, dass entweder

$$\frac{P(Y_1 = 0, Y_2 = 0, R_2 = 0)}{P(Y_1 = 1, Y_2 = 0, R_2 = 0)} \leq \frac{P(Y_1 = 0, R_2 = 1)}{P(Y_1 = 1, R_2 = 1)} \leq \frac{P(Y_1 = 0, Y_2 = 1, R_2 = 0)}{P(Y_1 = 1, Y_2 = 1, R_2 = 0)} \tag{6.23}$$

oder

$$\frac{P(Y_1 = 0, Y_2 = 0, R_2 = 0)}{P(Y_1 = 1, Y_2 = 0, R_2 = 0)} > \frac{P(Y_1 = 0, R_2 = 1)}{P(Y_1 = 1, R_2 = 1)} > \frac{P(Y_1 = 0, Y_2 = 1, R_2 = 0)}{P(Y_1 = 1, Y_2 = 1, R_2 = 0)} \tag{6.24}$$

gelten muss, da ansonsten die Restriktion

$$0 \leq P(Y_1 = a, Y_2 = b, R_2 = 1) \leq 1 \qquad \forall a, b = 0,1$$

verletzt ist.[96] Dementsprechend muss bei der Schätzung der Wahrscheinlichkeiten $P(Y_1 = a, Y_2 = b, R_2 = 1)$ beachtet werden, dass die Schätzer ebenfalls im zulässigen Parameterbereich liegen.[97]

Beispiel 6.3:

Zur Schätzung des Parameters θ der Verteilung $P_\theta(Y_1, Y_2)$ wird eine einfache Stichprobe vom Umfang n = 100 aus der Grundgesamtheit gezogen. Im Weiteren ist n_{ab}^R die Anzahl der vollständig beobachteten Untersuchungseinheiten mit $Y_1 = a$ und $Y_2 = b$ (a, b = 0, 1) und $n_{a\bullet}^{NR}$ die absolute Häufigkeit der Merkmalsträger mit $Y_1 = a$ und fehlendem Wert von Y_2 ($n_{a\bullet}^{NR} = n_{a0}^{NR} + n_{a1}^{NR}$). Die folgenden absoluten Kontingenztabellen umfassen – basierend auf obiger Einteilung – die Werte in der Stichprobe:

Y_1 und Y_2 beobachtet (R_2 = 0)

	Y_2=0	Y_2=1	
Y_1=0	$n_{00}^R = 16$	$n_{01}^R = 6$	$n_{0\bullet}^R = 22$
Y_1=1	$n_{10}^R = 32$	$n_{11}^R = 2$	$n_{1\bullet}^R = 34$
	$n_{\bullet 0}^R = 48$	$n_{\bullet 1}^R = 8$	$n^R = 56$

Y_2 nicht beobachtet (R_2 = 1)

	Y_2=0	Y_2=1	
Y_1=0	$n_{00}^{NR} = ?$	$n_{01}^{NR} = ?$	$n_{0\bullet}^{NR} = 28$
Y_1=1	$n_{10}^{NR} = ?$	$n_{11}^{NR} = ?$	$n_{1\bullet}^{NR} = 16$
	$n_{\bullet 0}^{NR} = ?$	$n_{\bullet 1}^{NR} = ?$	$n^{NR} = 44$

Allgemein sind die Parameter $(\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11})$ wegen

$$\theta_{ab} = P(Y_1 = a, Y_2 = b) = P(Y_1 = a, Y_2 = b, R_2 = 0) + P(Y_1 = a, Y_2 = b, R_2 = 1) \quad \forall a, b = 0,1$$

durch

$$\hat{\theta}_{ab} = \frac{n_{ab}^R}{n} + \frac{\hat{n}_{ab}^{NR}}{n} \tag{6.25}$$

zu schätzen, wobei $\hat{n}_{ab}^{NR}$ der Schätzer für die absolute Häufigkeiten n_{ab}^{NR} unter der jeweiligen Annahme (MCAR, MAR oder MNAR) ist. Falls der Ausfallmechanismus vom Typ MNAR ist und die bedingte Unabhängigkeitsbeziehung in (6.12) gilt, sind die in (6.19)-(6.22) hergeleiteten Gleichungen für $P(Y_1 = a, Y_2 = b, R_2 = 1)$

[96] Vgl. Baker/Laird (1988), S. 67.
[97] Auf diese Problematik wird im Beispiel 6.4 näher eingegangen.

(a, b = 0, 1) erfüllt. Durch Einsetzen der Schätzer in diese Gleichungen kann der Parametervektor $\theta = (\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11})$ in der folgenden Weise geschätzt werden:

$$\hat{\theta}_{00} = \frac{n_{00}^{R}}{n} + \frac{\hat{n}_{00}^{NR}}{n} = \frac{n_{00}^{R}}{n} + \frac{n_{1\bullet}^{NR} - \dfrac{n_{0\bullet}^{NR} n_{11}^{R}}{n_{01}^{R}}}{\dfrac{n_{10}^{R}}{n_{00}^{R}} - \dfrac{n_{11}^{R}}{n_{01}^{R}}} = 0{,}16 + 0{,}04 = 0{,}2 \tag{6.26}$$

$$\hat{\theta}_{10} = \frac{n_{10}^{R}}{n} + \frac{\hat{n}_{10}^{NR}}{n} = \frac{n_{10}^{R}}{n} + \frac{n_{1\bullet}^{NR} - \dfrac{n_{0\bullet}^{NR} n_{11}^{R}}{n_{01}^{R}}}{1 - \dfrac{n_{11}^{R} n_{00}^{R}}{n_{01}^{R} n_{10}^{R}}} = 0{,}32 + 0{,}08 = 0{,}4 \tag{6.27}$$

$$\hat{\theta}_{01} = \frac{n_{0\bullet}^{R}}{n} + \frac{n_{0\bullet}^{NR}}{n} - \hat{\theta}_{00} = 0{,}3 \tag{6.28}$$

$$\hat{\theta}_{11} = \frac{n_{1\bullet}^{R}}{n} + \frac{n_{1\bullet}^{NR}}{n} - \hat{\theta}_{10} = 0{,}1 \tag{6.29}$$

Wird hingegen von der Gültigkeit der bedingten Unabhängigkeitsbeziehung

$R_2 \perp\!\!\!\perp Y_2 \mid Y_1$

und somit von der Erfüllung der MAR-Annahme ausgegangen, so ist wegen

$$P(Y_2 = b \mid Y_1 = a, R_2 = 0) = P(Y_2 = b \mid Y_1 = a, R_2 = 1) \quad (\forall a, b = 0{,}1)$$

$$\Rightarrow \frac{P(Y_1 = a, Y_2 = b, R_2 = 0)}{P(Y_1 = a, R_2 = 0)} = \frac{P(Y_1 = a, Y_2 = b, R_2 = 1)}{P(Y_1 = a, R_2 = 1)}$$

die Gleichung

$$P(Y_1 = a, Y_2 = b, R_2 = 1) = \frac{P(Y_1 = a, Y_2 = b, R_2 = 0) P(Y_1 = a, R_2 = 1)}{P(Y_1 = a, R_2 = 0)}$$

erfüllt. Für den Schätzer in (6.25) gilt somit unter der MAR-Annahme

$$\hat{\theta}_{ab} = \frac{n_{ab}^{R}}{n} + \frac{\hat{n}_{ab}^{NR}}{n} = \frac{n_{ab}^{R}}{n} + \frac{n_{ab}^{R} n_{a\bullet}^{NR}}{n\, n_{a\bullet}^{R}} = \frac{n_{ab}^{R}}{n}\left(1 + \frac{n_{a\bullet}^{NR}}{n_{a\bullet}^{R}}\right),$$

so dass die Parameter unter einem ignorierbaren Ausfallmechanismus durch

$$\hat{\theta}_{00} = 0{,}364 \quad , \qquad \hat{\theta}_{10} = 0{,}471 \quad , \qquad \hat{\theta}_{01} = 0{,}136 \quad , \qquad \hat{\theta}_{11} = 0{,}029$$

zu schätzen sind. Anhand dieses Beispiels wird bereits deutlich, welche Konsequenzen sich für die Parameterschätzungen durch die unterschiedlichen Annahmen bezüglich des Ausfallmechanismus (MAR bzw. MNAR) ergeben können. Diese Auswirkungen werden durch eine Datensimulation, die anhand verschiedener Parameterwerte von $\psi = (\psi_1, \psi_2)$ durchgeführt wird, in Kapitel 7.2 näher untersucht.

Damit die geschätzten relativen Häufigkeiten $\frac{\hat{n}_{ab}^{NR}}{n}$ $(\forall a, b = 0{,}1)$ zulässige Lösungen im Sinne von

$$0 \le \frac{\hat{n}_{ab}^{NR}}{n} \le 1$$

bei Vorliegen des MNAR-Ausfallmechanismus sind, muss aufgrund von (6.23) und (6.24) entweder

$$\frac{n_{00}^{R}}{n_{10}^{R}} \le \frac{n_{0\bullet}^{NR}}{n_{1\bullet}^{NR}} \le \frac{n_{01}^{R}}{n_{11}^{R}} \tag{6.30}$$

oder

$$\frac{n_{00}^{R}}{n_{10}^{R}} > \frac{n_{0\bullet}^{NR}}{n_{1\bullet}^{NR}} > \frac{n_{01}^{R}}{n_{11}^{R}} \tag{6.31}$$

gelten.[98] Die erstgenannte Bedingung ist im vorangegangenen Beispiel mit

$$\frac{n_{00}^{R}}{n_{10}^{R}} = 0{,}5 \quad , \quad \frac{n_{0\bullet}^{NR}}{n_{1\bullet}^{NR}} = 1{,}75 \quad , \quad \frac{n_{01}^{R}}{n_{11}^{R}} = 3$$

erfüllt. Das folgende Beispiel zeigt dagegen, wie im anderen Fall der Nichteinhaltung von (6.30) oder (6.31) zu verfahren ist.

[98] Vgl. Baker/Laird (1988), S. 67.

Beispiel 6.4:

Es seien die gleichen absoluten Häufigkeiten n_{ab}^{R} $(a,b=0,1)$ wie im Beispiel 6.3 beobachtet worden (n = 100), die Randhäufigkeiten für Y_1 bei Nichtbeobachtung von Y_2 sind – abweichend von Beispiel 6.3 – durch $n_{0\bullet}^{NR}=34$ und $n_{1\bullet}^{NR}=10$ gegeben.

Y_1 und Y_2 beobachtet (R_2 = 0)

	Y_2=0	Y_2=1	
Y_1=0	$n_{00}^{R}=16$	$n_{01}^{R}=6$	$n_{0\bullet}^{R}=22$
Y_1=1	$n_{10}^{R}=32$	$n_{11}^{R}=2$	$n_{1\bullet}^{R}=34$
	$n_{\bullet 0}^{R}=48$	$n_{\bullet 1}^{R}=8$	$n^{R}=56$

Y_2 nicht beobachtet (R_2 = 1)

	Y_2=0	Y_2=1	
Y_1=0	$n_{00}^{NR}=?$	$n_{01}^{NR}=?$	$n_{0\bullet}^{NR}=34$
Y_1=1	$n_{10}^{NR}=?$	$n_{11}^{NR}=?$	$n_{1\bullet}^{NR}=10$
	$n_{\bullet 0}^{NR}=?$	$n_{\bullet 1}^{NR}=?$	$n^{NR}=44$

In diesem Beispiel gilt

$$\frac{n_{00}^{R}}{n_{10}^{R}} \le \frac{n_{01}^{R}}{n_{11}^{R}} \le \frac{n_{0\bullet}^{NR}}{n_{1\bullet}^{NR}},$$

und unter dem MNAR-Ausfallmechanismus würde die Schätzung von n_{00}^{NR} sowie n_{10}^{NR} durch

$$\hat{n}_{00}^{NR} = n\frac{n_{1\bullet}^{NR} - \dfrac{n_{0\bullet}^{NR} n_{11}^{R}}{n_{01}^{R}}}{\dfrac{n_{10}^{R}}{n_{00}^{R}} - \dfrac{n_{11}^{R}}{n_{01}^{R}}}$$

bzw.

$$\hat{n}_{10}^{NR} = n\frac{n_{1\bullet}^{NR} - \dfrac{n_{0\bullet}^{NR} n_{11}^{R}}{n_{01}^{R}}}{1 - \dfrac{n_{11}^{R} n_{00}^{R}}{n_{01}^{R} n_{10}^{R}}}$$

zu negativen und damit unzulässigen Werten führen.[99] Als Lösungen sind in dem Fall die kleinsten zulässigen Werte $\hat{n}_{00}^{NR} = 0$ und $\hat{n}_{10}^{NR} = 0$ zu verwenden,[100] so dass sich die folgenden Schätzer für die Parameter θ_{ab} ($a,b = 0,1$) ergeben:

$$\hat{\theta}_{00} = \frac{n_{00}^{R}}{n} + \frac{\hat{n}_{00}^{NR}}{n} = \frac{n_{00}^{R}}{n} = 0{,}16 \qquad \hat{\theta}_{10} = \frac{n_{10}^{R}}{n} + \frac{\hat{n}_{10}^{NR}}{n} = \frac{n_{10}^{R}}{n} = 0{,}32$$

$$\hat{\theta}_{01} = \frac{n_{0\bullet}^{R}}{n} + \frac{n_{0\bullet}^{NR}}{n} - \hat{\theta}_{00} = 0{,}4 \qquad \theta_{11} = \frac{n_{1\bullet}^{R}}{n} + \frac{n_{1\bullet}^{NR}}{n} - \hat{\theta}_{10} = 0{,}12$$

6.2.4 Fazit

In diesem Kapitel wurden zwei Verfahren vorgestellt, deren Ziel die Parameterschätzung einer diskreten Verteilung unter einem MNAR-Ausfallmechanismus ist. Um diese Schätzung zu ermöglichen, müssen – wie bei allen Behandlungsmethoden unter nicht ignorierbarem Datenausfall – entsprechende Annahmen erfüllt sein. Bei der Betrachtung einer diskreten Zufallsvariable (Kapitel 6.2.2) beziehen sich diese Annahmen auf eine Ratio von bedingten Ausfallwahrscheinlichkeiten, die den Wert des Parameterschätzers unmittelbar beeinflusst. Ein wesentlicher Vorteil des Verfahrens besteht in der Möglichkeit, die Sensitivität der Schätzung durch Variierung dieser Ratio bestimmen zu können.

Werden zwei binäre Zufallsvariablen betrachtet und weist lediglich eine der beiden Zufallsvariablen fehlende Werte in einer Stichprobe auf, kann das in Kapitel 6.2.3 beschriebene Verfahren angewendet werden, um die Verteilungsparameter bei Vorliegen eines Ausfallmechanismus vom Typ MNAR zu schätzen. Die Annahme, die in diesem Fall getroffen werden muss, ist die Gültigkeit einer bestimmten bedingten Unabhängigkeitsbeziehung.[101] In einem Beispiel wurde gezeigt, welchen Einfluss der unterstellte Ausfalltyp (MAR bzw. MNAR) auf die Parameterschätzungen besitzt.[102] Aus diesem Grund ist es sinnvoll, die Sensitivität der Schätzung zu beurteilen, indem

---

[99] In dieser Weise wurden die Schätzer im vorangegangenen Beispiel berechnet (vgl. Gleichungen (6.26) und (6.27)).
[100] Vgl. Baker/Laird (1988), S. 67.
[101] Vgl. Bedingung in (6.12).
[102] Vgl. Beispiel 6.3.

die Parameterschätzer unter den beiden Ausfalltypen ermittelt werden.

Werden mehr als zwei Zufallsvariablen betrachtet oder sind die den speziellen Verfahren zugrunde liegenden Annahmen nicht gerechtfertigt, so sind die in den folgenden beiden Kapiteln diskutierten Selection- und Pattern-Mixture Modelle zur Behandlung von fehlenden Werten anzuwenden. Die beiden Ansätze unterscheiden sich dabei in der Faktorisierung der gemeinsamen Verteilung $P(\mathbf{Y}, \mathbf{R})$. Innerhalb der Selection Modelle ist der Ausfallmechanismus $P_{\psi}(\mathbf{R} \mid \mathbf{y})$ zu modellieren, um den Parameter θ der Verteilung $P_{\theta}(\mathbf{Y})$ schätzen zu können. Die Erkenntnisse aus Kapitel 4 stehen im Einklang mit diesen Modellen,[103] so dass prinzipiell die likelihood-basierten Verfahren[104] zur Schätzung von θ angewendet werden können. Allerdings ist zu berücksichtigen, dass aufgrund der in diesem Kapitel behandelten, nicht ignorierbaren Ausfallmechanismen Inferenzen bezüglich θ aus der Likelihood-Funktion $L(\mathbf{r}, \mathbf{y}_{obs} \mid \theta, \psi)$ zu ziehen sind.

Pattern-Mixture Modelle stellen einen neuen und – im Vergleich zu Selection Modellen – grundlegend verschiedenen Ansatz zur Behandlung von fehlenden Werten dar. Bei diesem Ansatz ist die bedingte Verteilung $P_{\omega}(\underline{Y} \mid \underline{R} = \underline{r})$ für alle Werte, welche die Indikatorvariable $\underline{R}$ annehmen kann, zu modellieren. Diese bedingten Verteilungen, die durch den Parametervektor ω beschrieben sind, werden anschließend über die Verteilung der Indikatorvariable $P_{\varepsilon}(\underline{R})$ „gemischt“, um die Randverteilung $P_{\omega,\varepsilon}(\underline{Y})$ zu erhalten.[105] Ein ausführlicher Vergleich von Pattern-Mixture und Selection Modellen erfolgt in einem gesonderten Kapitel.[106]

---

[103] Insbesondere wurde an dieser Stelle nachgewiesen, dass bei Ignorierbarkeit des Ausfallmechanismus der Parameter θ aus der Likelihood-Funktion der beobachteten Daten $L(\mathbf{y}_{obs} \mid \theta)$ geschätzt werden kann (vgl. Herleitungen in (4.3) und (4.10)).

[104] Vgl. Kapitel 5.3.

[105] Vgl. Toutenburg et al. (2004), S. 30.

[106] Vgl. Kapitel 6.5.

6.3 Selection Modelle

6.3.1 Zielsetzung

Im Kontext der Problematik von fehlenden Werten sind Methoden von zentraler Bedeutung, die es ermöglichen, den Parametervektor θ der marginalen Verteilung von $\underline{Y}$ bei verschiedenen nicht ignorierbaren Ausfallmechanismen zu schätzen. Diese Schätzung bildet u.a. auch die Grundlage für eine geeignete Ersetzung der fehlenden Werte mit dem Ziel, ein vervollständigtes Datenmaterial zu erhalten und Standardverfahren zur statistischen Analyse anwenden zu können. Ein grundlegender, als Selection Modell bezeichneter Ansatz zur Bestimmung von θ beruht auf der folgenden Faktorisierung der gemeinsamen Verteilung $P_{\theta,\psi}(\mathbf{Y},\mathbf{R})$:[107]

$$P_{\theta,\psi}(\mathbf{Y},\mathbf{R}) = P_{\theta}(\mathbf{Y})\, P_{\psi}(\mathbf{R} \mid \mathbf{Y}) \tag{6.32}$$

Innerhalb eines Selection Modells wird die Verteilung von **Y** separat vom Ausfallmechanismus beschrieben. Dies impliziert, dass sowohl die beobachteten Daten $\mathbf{y}_{obs}$ als auch die fehlenden Werte $\mathbf{y}_{mis}$ durch eine Verteilung mit dem Parametervektor θ modelliert werden können.

Die Verteilung des Ausfallmechanismus **R** in (6.32) ist von **y** abhängig, so dass die (Nicht-)Beobachtung einer Variable durch die Werte in **y** bestimmt wird. Aus dieser Eigenschaft resultiert die Bezeichnung „Selection Modelle“, da durch den zugrunde liegenden Ausfallmechanismus die Selektion der beobachteten Daten $\mathbf{y}_{obs}$ erfolgt.[108]

Ist der Datenausfall lediglich auf die beobachteten Daten $\mathbf{y}_{obs}$ zurückzuführen, so kann $P_{\psi}(\mathbf{R} \mid \mathbf{y}_{obs}, \mathbf{y}_{mis})$ aufgrund von

$$P_{\psi}(\mathbf{R} \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) = P_{\psi}(\mathbf{R} \mid \mathbf{y}_{obs})$$

für die Schätzung von θ vernachlässigt bzw. ignoriert werden.[109] Im Fall von MNAR ist der Ausfallmechanismus jedoch zu berücksichtigen, wodurch die Schätzung von θ nur unter Annahmen bezüglich der Verteilung $P_{\psi}(\mathbf{R} \mid \mathbf{y})$ möglich ist. Diese Annahmen betreffen insbesondere den Verteilungstyp, über den aus inhaltlichen Gründen

---

[107] Vgl. Herleitungen in (4.3) für die frequentistische und (4.10) für die bayesianische Sichtweise.
[108] Vgl. Rubin (1987), S. 207.
[109] Vgl. Formel (4.3).

allenfalls grobe Anhaltspunkte in der praktischen Anwendung vorliegen. Hierdurch wird zusätzlich eine Sensitivitätsanalyse bezüglich des Ausfallmechanismus erschwert, da sich eine Variierung des Verteilungstyps und die jeweils nachfolgende Parameterbestimmung als kompliziert und aufwendig erweist. Um den Parametervektor θ schätzen zu können, wird außerdem die Kenntnis des Verteilungstyps von $\underline{Y}$ vorausgesetzt.

Das Ziel der folgenden Ausführungen ist es, die Methodik der Selection Modelle am Beispiel einer normalverteilten Zufallsvariable Y_1 zu verdeutlichen. Es wird gezeigt, wie unter verschiedenen Modellierungen (Schwellenwert-Modellierung, Logit-Modellierung) des nicht ignorierbaren Ausfallmechanismus eine Schätzung der Parameter der Normalverteilung von Y_1 erfolgen kann. Für eine Erweiterung dieses Modells, bei der eine ebenfalls normalverteilte Kovariate Y_2 zur Parameterschätzung einbezogen wird, sei auf Little/Rubin (2002) verwiesen.[110]

6.3.2 Univariates Selection Modell bei Normalverteilung

6.3.2.1 Vorbetrachtung

Im Mittelpunkt dieses Kapitels steht die Schätzung des Parametervektors θ der Verteilung $P_\theta(Y_1)$ einer Zufallsvariable Y_1, zu deren Zweck eine einfache Stichprobe vom Umfang *n* aus der Grundgesamtheit entnommen wurde. Die Zufallsvariable Y_1 ist dabei normalverteilt mit Erwartungswert μ und Varianz σ^2 $(\theta = (\mu, \sigma))$:

$$Y_1 \sim N(\mu, \sigma^2)$$

Ferner sind in der Stichprobe *q* Realisationen von Y_1 beobachtet und (*n*-*q*) Werte nicht beobachtet worden.[111] Der Datenausfall hängt dabei von den Realisationen der Zufallsvariable Y_1 ab, so dass der Ausfallmechanismus vom Typ MNAR ist. Da der Verteilungstyp von Y_1 annahmegemäß bekannt ist, kann eine Selection Modellierung unter Spezifizierung des vorliegenden, nicht ignorierbaren Ausfallmechanismus

---

[110] Vgl. Little/Rubin (2002), S. 322f. Little/Rubin beschränken sich dabei auf die Schwellenwert-Modellierung des Ausfallmechanismus.

[111] O.B.d.A. werden die Daten in der Weise geordnet, dass die ersten *q* Werte $y_{11},\ldots,y_{q1}$ beobachtet worden sind.

durchgeführt werden. Ausgehend von dem Selection Modell in (6.32) gilt für $P_{\theta,\psi}(\mathbf{y},\mathbf{r})$:

$$\begin{aligned} P_{\theta,\psi}(\mathbf{y},\mathbf{r}) &= \prod_{i=1}^{n} P_{\theta,\psi}(y_{i1},r_{i1}) = \prod_{i=1}^{n} P_{\theta}(y_{i1})P_{\psi}(r_{i1}\mid y_{i1}) \\ &= \prod_{i=1}^{q} P_{\theta}(y_{i1})P_{\psi}(R_{i1}=0\mid y_{i1}) \prod_{i=q+1}^{n} P_{\theta}(y_{i1})P_{\psi}(R_{i1}=1\mid y_{i1}) \\ &= \prod_{i=1}^{q} P_{\theta}(\underline{y}_{obs,i})P_{\psi}(R_{i1}=0\mid \underline{y}_{obs,i}) \prod_{i=q+1}^{n} P_{\theta}(\underline{y}_{mis,i})P_{\psi}(R_{i1}=1\mid \underline{y}_{mis,i}) \end{aligned} \tag{6.33}$$

Durch Integration von $P_{\theta,\psi}(\mathbf{y},\mathbf{r})$ über die fehlenden Werte $\mathbf{y}_{mis} = (\underline{y}_{mis,q+1},...,\underline{y}_{mis,n})$ erhält man $P_{\theta,\psi}(\mathbf{y}_{obs},\mathbf{r})$:

$$P_{\theta,\psi}(\mathbf{y}_{obs},\mathbf{r}) = \prod_{i=1}^{q} P_{\theta}(\underline{y}_{obs,i})P_{\psi}(R_{i1}=0\mid \underline{y}_{obs,i}) \prod_{i=q+1}^{n} \int_{-\infty}^{\infty} P_{\theta}(\underline{y}_{mis,i})P_{\psi}(R_{i1}=1\mid \underline{y}_{mis,i})\, d\underline{y}_{mis,i} \tag{6.34}$$

Für die Dichtefunktion $P_{\theta}(y_{i1})$ mit $\theta=(\mu,\sigma)$ gilt[112]

$$P_{\theta}(y_{i1}) = \phi\left(\frac{y_{i1}-\mu}{\sigma}\right)\frac{1}{\sigma} \qquad (i = 1,\ldots,n), \tag{6.35}$$

wobei $\phi(\cdot)$ die Dichte der Standardnormalverteilung ist.

Ausgehend von (6.34) und (6.35) gilt für die Likelihood-Funktion $L(\mathbf{y}_{obs},\mathbf{r}\mid\mu,\sigma,\psi)$:

$$\begin{aligned} L(\mathbf{y}_{obs},\mathbf{r}\mid\mu,\sigma,\psi) = &\prod_{i=1}^{q} \phi\left(\frac{\underline{y}_{obs,i}-\mu}{\sigma}\right)\frac{1}{\sigma} P_{\psi}(R_{i1}=0\mid \underline{y}_{obs,i}) \\ &\cdot \prod_{i=q+1}^{n} \int_{-\infty}^{\infty} \phi\left(\frac{\underline{y}_{mis,i}-\mu}{\sigma}\right)\frac{1}{\sigma} P_{\psi}(R_{i1}=1\mid \underline{y}_{mis,i})\, d\underline{y}_{mis,i} \end{aligned} \tag{6.36}$$

Um den interessierenden Parametervektor $\theta=(\mu,\sigma)$ aus dieser Likelihood-Funktion schätzen zu können, ist die Modellierung des nicht ignorierbaren Ausfallmechanismus

$$P_{\psi}(\mathbf{R}\mid\mathbf{y}) = \prod_{i=1}^{n} P_{\psi}(R_{i1}=r_{i1}\mid Y_{i1}=y_{i1})$$

[112] Vgl. Greene (2000), S. 66.

notwendig. Selbst unter dieser Spezifizierung des Ausfallmechanismus erweist sich die Maximierung der Likelihood-Funktion $L(\mathbf{y}_{obs}, \mathbf{r} \mid \theta, \psi)$ bezüglich θ im Allgemeinen als schwierig und kann häufig nur durch iterative Methoden (z.B. den Newton-Raphson Algorithmus, EM-Algorithmus) erfolgen.[113] Im Folgenden soll diese Problematik anhand von zwei verschiedenen Modellen (Schwellenwert-Modell, Logit-Modell) des Ausfallmechanismus verdeutlicht werden.

6.3.2.2 Schwellenwert-Modellierung des Ausfallmechanismus

Das in diesem Kapitel betrachtete Modell für den Ausfallmechanismus ist durch die folgenden Ausfallwahrscheinlichkeiten charakterisiert:

$$P_\psi(R_1 = 1 \mid Y_1 \le 0) = \psi$$
$$P_\psi(R_1 = 1 \mid Y_1 > 0) = 0 \tag{6.37}$$

Im Weiteren wird davon ausgegangen, dass q_1 Werte mit $y_{i1} \le 0$ ($i = 1,\ldots,q_1$) beobachtet wurden ($q_1 < q$). Dann gilt für die Likelihood-Funktion $L(\mathbf{y}_{obs}, \mathbf{r} \mid \mu, \sigma, \psi)$ in (6.36):

$$\begin{aligned}
L(\mathbf{y}_{obs}, \mathbf{r} \mid \mu, \sigma, \psi) &= \prod_{i=1}^{q_1} \phi\left(\frac{\underline{y}_{obs,i} - \mu}{\sigma}\right)\frac{1}{\sigma} P_\psi(R_{i1} = 0 \mid Y_{i1} \le 0) \\
&\quad \cdot \prod_{i=q_1+1}^{q} \phi\left(\frac{\underline{y}_{obs,i} - \mu}{\sigma}\right)\frac{1}{\sigma} P_\psi(R_{i1} = 0 \mid Y_{i1} > 0) \\
&\quad \cdot \prod_{i=q+1}^{n} \int_{-\infty}^{0} \phi\left(\frac{\underline{y}_{mis,i} - \mu}{\sigma}\right)\frac{1}{\sigma} P_\psi(R_{i1} = 1 \mid Y_{i1} \le 0) d\,\underline{y}_{mis,i} \\
&= \prod_{i=1}^{q_1} \phi\left(\frac{\underline{y}_{obs,i} - \mu}{\sigma}\right)\frac{1}{\sigma}(1-\psi) \prod_{i=q_1+1}^{q} \phi\left(\frac{\underline{y}_{obs,i} - \mu}{\sigma}\right)\frac{1}{\sigma} \\
&\quad \cdot \prod_{i=q+1}^{n} \Phi\left(\frac{0-\mu}{\sigma}\right)\frac{1}{\sigma}\psi \\
&= (1-\psi)^{q_1}\psi^{(n-q)}\frac{1}{\sigma^n}\prod_{i=1}^{q} \phi\left(\frac{\underline{y}_{obs,i} - \mu}{\sigma}\right)\prod_{i=q+1}^{n} \Phi\left(\frac{0-\mu}{\sigma}\right)
\end{aligned} \tag{6.38}$$

[113] Vgl. Little (1994), S. 474.

Im Spezialfall $\psi = 1$ werden alle Werte $y_1 \leq 0$ in der Stichprobe nicht beobachtet und man erhält ein zensiertes Tobit-Modell.[114] In diesem Tobit-Modell sowie im allgemeinen Fall mit $0 < \psi < 1$ können die Parameter μ und σ nicht direkt aus der Likelihood-Funktion geschätzt werden, so dass im Folgenden der EM-Algorithmus als eine iterative Methode zur Bestimmung von $\hat{\mu}$ und $\hat{\sigma}$ angewendet wird.

Für die Schätzung des Parametervektors $\theta = (\mu, \sigma)$ ist es ausreichend, die suffizienten Statistiken der Normalverteilung $\sum_i^n y_{i1}$ und $\sum_i^n y_{i1}^2$ zu betrachten.[115] Da (*n*-*q*) Werte in der Stichprobe nicht beobachtet wurden, ist im ersten Schritt des EM-Algorithmus der bedingte Erwartungswert von $\sum_{i=1}^n Y_{i1}$ gegeben die Ausgangswerte $\theta^{(0)} = \left(\mu^{(0)}, \sigma^{(0)}\right)$ und $\psi^{(0)}$ zu bilden:[116]

$$E\left(\sum_{i=1}^n Y_{i1} \mid \theta^{(0)}, \psi^{(0)}, \mathbf{y}_{obs}, \mathbf{r}\right) = \sum_{i=1}^q y_{i1} + E\left(\sum_{i=q+1}^n Y_{i1} \mid \theta^{(0)}, \psi^{(0)}, \mathbf{y}_{obs}, \mathbf{r}\right)$$
$$= \sum_{i=1}^q y_{i1} + (n-q) E\left(Y_1 \mid \theta^{(0)}, \psi^{(0)}, R_1 = 1\right) \qquad (6.39)$$

Für den bedingten Erwartungswert $E\left(Y_1 \mid \theta^{(0)}, \psi^{(0)}, R_1 = 1\right)$ gilt:[117]

$$E\left(Y_1 \mid \theta^{(0)}, \psi^{(0)}, R_1 = 1\right) = E(Y_1 \mid \theta^{(0)}, \psi^{(0)}, Y_1 \leq 0)$$
$$= \mu^{(0)} + \sigma^{(0)} \tilde{\lambda}\left(\frac{0 - \mu^{(0)}}{\sigma^{(0)}}\right) \qquad ; \tilde{\lambda}(z) = \frac{-\phi(z)}{\Phi(z)} \qquad (6.40)$$

Analog wird der bedingte Erwartungswert der Statistik $\sum_i^n Y_{i1}^2$ gegeben dem Parametervektor $\theta^{(0)}$ und $\psi^{(0)}$ berechnet:[118]

[114] Vgl. Ronning (1991), S. 124ff.
[115] Vgl. Bamberg (1972), S. 71.
[116] Vgl. Little/Rubin (2002), S. 168.
[117] Vgl. Greene (2000), S. 899.
[118] Vgl. Little/Rubin (2002), S. 168.

$$E\left(\sum_{i=1}^{n} Y_{i1}^2 \mid \theta^{(0)}, \psi^{(0)}, \mathbf{y}_{obs}, \mathbf{r}\right) = \sum_{i=1}^{q} y_{i1}^2 + (n-q)E\left(Y_1^2 \mid \theta^{(0)}, \psi^{(0)}, R_1 = 1\right)$$
$$= \sum_{i=1}^{q} y_{i1}^2 + (n-q)\left[\left(E\left(Y_1 \mid \theta^{(0)}, \psi^{(0)}, R_1 = 1\right)\right)^2 + \mathrm{Var}(Y_1 \mid \theta^{(0)}, \psi^{(0)}, R_1 = 1)\right] \quad (6.41)$$

Für das Moment $\mathrm{Var}(Y_1 \mid \theta^{(0)}, \psi^{(0)}, R_1 = 1)$ in (6.41) gilt:[119]

$$\mathrm{Var}(Y_1 \mid \theta^{(0)}, \psi^{(0)}, R_1 = 1) = \mathrm{Var}(Y_1 \mid \theta^{(0)}, \psi^{(0)}, Y_1 \le 0)$$
$$= \left(\sigma^{(0)}\right)^2\left(1 - \tilde{\delta}\left(\frac{0-\mu^{(0)}}{\sigma^{(0)}}\right)\right) \quad ; \tilde{\delta}(z) = \tilde{\lambda}(z)\left(\tilde{\lambda}(z) - z\right) \quad (6.42)$$

Die bedingten Erwartungswerte der suffizienten Statistiken hängen nicht von dem Parameter $\psi^{(0)}$ ab. Aus diesem Grund kann im Folgenden auf die Bestimmung von $\psi^{(t)}$ in Iteration *t* verzichtet werden.

Im M-Schritt sind die Parameter $\mu^{(1)}$ und $\sigma^{(1)}$ zu ermitteln, wobei die unbeobachteten Daten durch den bedingten Erwartungswert in (6.40) ersetzt werden:[120]

$$\mu^{(1)} = \frac{1}{n} E\left(\sum_{i=1}^{n} Y_{i1} \mid \theta^{(0)}, \mathbf{y}_{obs}, \mathbf{r}\right) = \frac{1}{n}\left[\sum_{i=1}^{q} y_{i1} + (n-q)E\left(Y_1 \mid \theta^{(0)}, R_1 = 1\right)\right] \quad (6.43)$$

$$\left(\sigma^{(1)}\right)^2 = \frac{1}{n} E\left(\sum_{i=1}^{n} Y_{i1}^2 \mid \theta^{(0)}, \mathbf{y}_{obs}, \mathbf{r}\right) - \left(\frac{1}{n} E\left(\sum_{i=1}^{n} Y_{i1} \mid \theta^{(0)}, \mathbf{y}_{obs}, \mathbf{r}\right)\right)^2$$
$$= \frac{1}{n}\sum_{i=1}^{q} y_{i1}^2 + \frac{(n-q)}{n} E\left(Y_1^2 \mid \theta^{(0)}, R_1 = 1\right) - \left(\mu^{(1)}\right)^2$$
$$= \frac{1}{n}\sum_{i=1}^{q} y_{i1}^2 + \frac{(n-q)}{n}\left[\left(E\left(Y_1 \mid \theta^{(0)}, R_1 = 1\right)\right)^2 + \mathrm{Var}(Y_1 \mid \theta^{(0)}, R_1 = 1)\right] - \left(\mu^{(1)}\right)^2$$
$$= \frac{1}{n}\sum_{i=1}^{q} y_{i1}^2 + \frac{(n-q)}{n}\left[\left(E\left(Y_1 \mid \theta^{(0)}, R_1 = 1\right)\right)^2 + \left(\sigma^{(0)}\right)^2\left(1 - \tilde{\delta}\left(\frac{0-\mu^{(0)}}{\sigma^{(0)}}\right)\right)\right] - \left(\mu^{(1)}\right)^2 \quad (6.44)$$

Ausgehend von diesen beiden Parametern wird der bedingte Erwartungswert $E\left(Y_1 \mid \theta^{(1)}, R_1 = 1\right)$ erneut bestimmt und anschließend $\theta^{(2)} = \left(\mu^{(2)}, \sigma^{(2)}\right)$ in analoger Weise zu (6.43) und (6.44) ermittelt. Das Verfahren wird wiederholt, bis die Parameter der einzelnen Iterationen konvergieren $\left(\theta^{(t)} \approx \theta^{(t-1)}\right)$.

---

[119] Vgl. Greene (2000), S. 899.
[120] Vgl. Little/Rubin (2002), S. 168.

Beispiel 6.5:

Von einer univariat normalverteilten Zufallsvariable Y_1 seien die folgenden Realisationen beobachtet worden:

i	Y_{i1}
1	-0,86
2	0,21
3	0,72
4	3,04
5	1,51
6	0,28
7	1,41
8	1,12
9	-0,17
10	-0,27

i	Y_{i1}
11	0,19
12	0,71
13	1,52
14	0,43
15	1,19
16	0,11
17	?
18	?
19	?
20	?

Weiterhin sei der folgende nicht ignorierbare Ausfallmechanismus bekannt:

$$P_\psi\left(R_{i1}=1 \mid Y_{i1} \le 0\right)=\psi$$
$$P_\psi\left(R_{i1}=1 \mid Y_{i1} > 0\right)=0 \qquad (i=1,...,n) \tag{6.45}$$

Als Startwerte für $\mu^{(0)}$ und $\sigma^{(0)}$ wurden die Parameter der Standardnormalverteilung gewählt:

$$\mu^{(0)}=0 \ , \ \sigma^{(0)}=1$$

Als Abbruchkriterium des EM-Algorithmus sei

$$\left|\mu^{(t+1)}-\mu^{(t)}\right| \le 0{,}0001 \quad , \quad \left|\sigma^{(t+1)}-\sigma^{(t)}\right| \le 0{,}0001$$

festgelegt.

Die folgende Tabelle umfasst die in Iteration *t* des Algorithmus ermittelten Werte für den Parametervektor $\theta^{(t)}=\left(\mu^{(t)},\sigma^{(t)}\right)$ sowie für $E\left(Y_1 \mid \theta^{(t)}, R_1=1\right)$.

t	$\mu^{(t)}$	$\sigma^{(t)}$	$E\left(Y_1 \mid \theta^{(t)}, R_1 = 1\right)$	$\left\|\mu^{(t+1)} - \mu^{(t)}\right\|$	$\left\|\sigma^{(t+1)} - \sigma^{(t)}\right\|$
0	0	1	-0,797885	0,397423	0,025197
1	0,397423	0,974803	-0,676839	0,024209	0,054910
2	0,421632	1,029713	-0,653897	0,004589	0,020537
3	0,426221	1,050250	-0,646704	0,001438	0,006213
4	0,427659	1,056463	-0,644532	0,000435	0,001908
5	0,428094	1,058371	-0,643869	0,000132	0,000584
6	0,428226	1,058955	-0,643666	0,000041	0,000179
7	0,428267	1,059134	-0,643604	0,000012	0,000054
8	0,428279	1,059188	-0,643585		

Tabelle 6.1: Parameterwerte und bedingter Erwartungswert von Y_1 innerhalb der einzelnen Iterationen des EM-Algorithmus

Die Lösungen des Algorithmus konvergieren im Beispiel bereits nach der 7. Iteration, und somit erhält man die Schätzwerte

$$\hat{\mu} = \mu^{(7)} = 0{,}428\,,$$
$$\hat{\sigma} = \sigma^{(7)} = 1{,}059 \tag{6.46}$$

für μ und σ. Diese können mit den unter Gültigkeit der MAR-Annahme resultierenden Schätzwerten

$$\hat{\mu} = \frac{1}{q}\sum_{i=1}^{q} y_{i1} = 0{,}7\,,$$
$$\hat{\sigma} = \sqrt{\frac{1}{q}\sum_{i=1}^{q} y_{i1}^2 - \hat{\mu}^2} = 0{,}9 \tag{6.47}$$

verglichen werden.[121] Der niedrigere Wert für $\hat{\mu}$ in (6.46) ist auf den MNAR-Ausfallmechanismus (6.45) zurückzuführen, da unter diesen Bedingungen lediglich niedrige Werte von Y_1 fehlen und der Schätzer in (6.47) nach oben verzerrt ist. Aus

[121] Da in diesem Kapitel lediglich eine Zufallsvariable betrachtet wird, ist dieser mit dem Schätzwert unter der MCAR-Annahme identisch.

dem gleichen Grund resultiert eine höhere geschätzte Varianz, wenn der Ausfallmechanismus – wie in (6.45) angenommen – nicht ignorierbar ist.

6.3.2.3 Logit-Modellierung des Ausfallmechanismus

Ist der Datenausfall einer normalverteilten Zufallsvariable Y_1 nicht von einem Schwellenwert abhängig, sind alternative Modelle zu (6.37) für den Ausfallmechanismus zu berücksichtigen. Das Logit-Modell

$$P_{\alpha,\beta}(R_1 = 1 \mid y_1) = \frac{1}{1 + e^{\left(-\frac{y_1 - \alpha}{\beta}\right)}} \tag{6.48}$$

mit $\psi = (\alpha, \beta)$, $\alpha = \mu$ und $\beta = \frac{\sqrt{3}}{\pi}\sigma$ stellt eine Möglichkeit dar, um die bedingte Ausfallwahrscheinlichkeit $P_{\alpha,\beta}(R_1 = 1 \mid y_1)$ zu beschreiben.[122, 123] Die Likelihood-Funktion für die Parameter μ, σ und ψ gegeben die beobachteten Daten ist dann[124]

$$L(\mathbf{y}_{obs}, \mathbf{r} \mid \mu, \sigma, \alpha, \beta) = \prod_{i=1}^{q} \phi\left(\frac{\underline{y}_{obs,i} - \mu}{\sigma}\right)\frac{1}{\sigma} P_{\alpha,\beta}(R_{i1} = 0 \mid \underline{y}_{obs,i})$$
$$\cdot \prod_{i=q+1}^{n} \int_{-\infty}^{\infty} \phi\left(\frac{\underline{y}_{mis,i} - \mu}{\sigma}\right)\frac{1}{\sigma} P_{\alpha,\beta}(R_{i1} = 1 \mid \underline{y}_{mis,i}) d\underline{y}_{mis,i}$$

$$= \prod_{i=1}^{q} \phi\left(\frac{\underline{y}_{i,obs} - \mu}{\sigma}\right)\frac{1}{\sigma} \frac{e^{\left(-\frac{\underline{y}_{i,obs} - \alpha}{\beta}\right)}}{1 + e^{\left(-\frac{\underline{y}_{i,obs} - \alpha}{\beta}\right)}} \prod_{i=q+1}^{n} \int_{-\infty}^{\infty} \phi\left(\frac{\underline{y}_{i,mis} - \mu}{\sigma}\right)\frac{1}{\sigma} \frac{1}{1 + e^{\left(-\frac{\underline{y}_{i,mis} - \alpha}{\beta}\right)}} d\underline{y}_{i,mis} . \tag{6.49}$$

[122] Vgl. Greene (2000), S. 860.
[123] Vgl. Voß (2000), S. 613.
[124] Vgl. Herleitung in Formel (6.36).

Im Gegensatz zu dem Schwellenwert-Modell kann der bedingte Erwartungswert $E(Y_1 \mid \mu, \sigma, \alpha, \beta, R_1 = 1)$ nicht direkt aus

$$E(Y_1 \mid \mu, \sigma, \alpha, \beta, R_1 = 1) = \frac{\int_{-\infty}^{\infty} y_1 \phi\left(\frac{y_1 - \mu}{\sigma}\right) \frac{1}{\sigma} \frac{1}{1 + e^{\left(-\frac{y_1 - \alpha}{\beta}\right)}} dy_1}{\int_{-\infty}^{\infty} \phi\left(\frac{y_1 - \mu}{\sigma}\right) \frac{1}{\sigma} \frac{1}{1 + e^{\left(-\frac{y_1 - \alpha}{\beta}\right)}} dy_1}$$

bestimmt werden, so dass eine andere Vorgehensweise als im vorigen Kapitel notwendig ist. Konkret ist der E-Schritt des EM-Algorithmus in der Weise zu modifizieren, dass – anstatt der Bildung von bedingten Erwartungswerten – Realisationen aus der Verteilung $P_{\alpha,\beta,\mu,\sigma}(Y_1 = y_1, R_1 = 1)$ für die nicht beobachteten Daten gezogen werden. Weiterhin werden die fehlenden Werte mehrfach ersetzt (Multiple Imputation mit $m = 5$), um die Variabilität bezüglich der Imputationen zu berücksichtigen.[125] Der folgende Ablaufplan verdeutlicht in Kurzform die Vorgehensweise innerhalb des veränderten Algorithmus, beginnend mit dem Datenbestand $h = 1$ $(1 \le h \le m)$:

1. Festlegung der Startwerte $\mu^{(h,0)}$, $\sigma^{(h,0)}$, $\alpha^{(h,0)}$ und $\beta^{(h,0)}$ ($t = 0$, $i = q + 1$)
2. Ziehen einen Wertes $y_{i1}^{(h,t)}$ aus der Normalverteilung $P\left(Y_{i1}^{(h,t)} \mid \mu^{(h,t)}, \sigma^{(h,t)}\right)$
3. Bestimmung von $P\left(R_{i1}^{(h,t)} = 1 \mid y_{i1}^{(h,t)}, \alpha^{(h,t)}, \beta^{(h,t)}\right) = \frac{1}{1 + e^{\left(-\frac{y_{i1}^{(h,t)} - \alpha^{(h,t)}}{\beta^{(h,t)}}\right)}}$
4. Für $P\left(R_{i1}^{(h,t)} = 1 \mid y_{i1}^{(h,t)}, \alpha^{(h,t)}, \beta^{(h,t)}\right) > Z_i^{(h,t)}$ mit der rechteckverteilten Zufallsvariable

   $Z_i^{(h,t)} \sim \mathrm{Re}(0;1)$

   wird der fehlende Wert durch $y_{i1}^{(h,t)}$ ersetzt, ansonsten wird das Verfahren beginnend mit Schritt 2 wiederholt
5. Durchführung der Schritte 2-4 für $i = (q + 2, ..., n)$

[125] Vgl. Kapitel 5.2.

6. Ermittlung der Parameterwerte für die nächste Iteration (t + 1):

$$\mu^{(h,t+1)} = \frac{1}{n}\sum_{i=1}^{n} y_{i1}^{(h,t)} \quad , \quad \sigma^{(h,t+1)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{i1}^{(h,t)} - \bar{y}_1^{(h,t)}\right)^2} \quad , \quad \alpha^{(h,t+1)} = \mu^{(h,t+1)} \quad ,$$

$$\beta^{(h,t+1)} = \frac{\sqrt{3}}{\pi}\sigma^{(h,t+1)}$$

7. Durchführung der Schritte 2-6 mit den neu ermittelten Parameterwerten bis zur Konvergenz der Parameterlösungen

8. Ausführung der Schritte 1-7 für $h = 2,\ldots,m$ (Multiple Imputation)

Beispiel 6.6:

Aus der Grundgesamtheit einer normalverteilten Zufallsvariable Y_1 mit $\mu = 0{,}5$ und $\sigma = 1$ wird eine Stichprobe vom Umfang $n = 5000$ gezogen. Innerhalb der Stichprobe wurden fehlende Werte durch das Logit-Modell in (6.48) mit $\alpha = 0{,}5$ und $\beta = 0{,}5513$ erzeugt, wodurch $q = 2485$ Werte mit Mittelwert

$$\frac{1}{q}\sum_{i=1}^{q} y_{i1} = 1{,}077$$

und Standardabweichung

$$\sqrt{\frac{1}{q}\sum_{i=1}^{q} {y_{i1}}^2 - \left(\frac{1}{q}\sum_{i=1}^{q} y_{i1}\right)^2} = 0{,}8289$$

beobachtet wurden. Als Startwerte des Algorithmus sind

$$\mu^{(h,0)} = 0,\ \sigma^{(h,0)} = 2,\ \alpha^{(h,0)} = 0, \beta^{(h,0)} = 1{,}1 \qquad \text{für } h = 1,\ldots,5$$

gewählt worden.

Die folgende Tabelle umfasst die Schätzwerte für μ und σ, die sich aus der beschriebenen multiplen Imputation der fehlenden Werte in den m Datenbeständen ergeben.

h	$\hat{\mu}^{(h)}$	$\hat{\sigma}^{(h)}$
1	0,4720	1,0271
2	0,4918	1,0057
3	0,4918	1,0068
4	0,4806	1,0291
5	0,4713	1,0364

Tabelle 6.2: Schätzwerte für μ und σ aus dem vervollständigten Datenbestand h

Die Schätzwerte für μ und σ sind somit:

$$\hat{\mu} = \frac{1}{5}\sum_{h=1}^{5}\hat{\mu}^{(h)} = 0{,}4815$$

$$\hat{\sigma} = \frac{1}{5}\sum_{h=1}^{5}\hat{\sigma}^{(h)} = 1{,}021$$

Für die Varianz des Schätzers $\hat{\mu}$ innerhalb eines Datenbestandes gilt

$$\overline{U} = \frac{1}{m}\sum_{h=1}^{m}U^{(h)} = \frac{1}{5}\sum_{h=1}^{5}\frac{\left(\hat{\sigma}^{(h)}\right)^2}{5000} = 0{,}0002\,,$$

und die Varianz zwischen den Imputationen ist

$$B = \frac{1}{m-1}\sum_{h=1}^{m}\left(\hat{\mu}^{(h)} - \hat{\mu}\right)^2 = 0{,}0001\,.$$[126]

Ausgehend von dem Schätzwert

$$T = \overline{U} + \left(1 + \frac{1}{m}\right)B = 0{,}0003$$

für die gesamte Varianz von $\hat{\mu}$ ist

$$\left[\hat{\mu} - t_{0,975}^{df}\sqrt{T}\;;\;\hat{\mu} + t_{0,975}^{df}\sqrt{T}\right] = \left[0{,}446\;;\;0{,}517\right]$$

[126] Vgl. Formeln (5.17) und (5.20).

ein 95%-Konfidenzintervall für den unbekannten Parameter μ.[127]

Aus

$$\hat{\beta} = \frac{\sqrt{3}}{\pi}\hat{\sigma} = 0{,}563 > 0$$

ist ersichtlich, dass hohe Werte von Y_1 häufiger beobachtet werden als niedrige Realisationen. Der Algorithmus ermöglicht dennoch unverzerrte Parameterschätzungen, indem der zugrunde liegende Ausfallmechanismus korrekt bestimmt wird und die fehlenden Werte vornehmlich durch niedrige Ausprägungen von Y_1 ersetzt werden. Obwohl keine Parameterannahmen getroffen werden bzw. auch kein Schwellenwert wie in (6.45) vorzugeben ist, sind die Schätzungen von Parametern in diesem Selection Modell in hohem Maße von den Verteilungsannahmen abhängig. Dies betrifft zum einen die Annahme einer Normalverteilung der *gesamten* Daten, die aufgrund der Nichtbeobachtung von einzelnen Merkmalsträgern nicht überprüft werden kann. Weiterhin wird von einem Logit-Ausfallmodell ausgegangen, eine Abgrenzung von anderen Modellen ist jedoch ebenfalls anhand der beobachteten Daten nicht möglich. Aus diesem Grund sind weitere Verteilungstypen sowohl für Y_1 als auch für den Ausfallmechanismus zu untersuchen, wobei die Sensitivitätsanalyse mit Hilfe des in diesem Kapitel beschriebenen Algorithmus durchgeführt werden kann.

6.3.3 Fazit

Die Selection Modellierung beruht auf der Zielstellung, einen interessierenden Parametervektor θ der Verteilung von $\underline{Y}$ aus den beobachteten Daten zu schätzen. Insbesondere kann durch diesen Ansatz die nahe liegende Betrachtungsweise, dass die (Nicht-)Beobachtung von Werten auf die Ausprägungen in der Datenmatrix **y** zurückzuführen ist, durch einen Ausfallmechanismus $P_{\psi}(\mathbf{R} \mid \mathbf{y})$ modelliert werden. Insofern ist auch die von Rubin definierte "Ignorability"-Eigenschaft des Ausfallmechanismus mit diesem Modell vereinbar, bei deren Erfüllung die in Kapitel 5.3 erwähnten likelihood-basierten Methoden zur Anwendung kommen können. Ist der

[127] Vgl. zur Berechnung der Freiheitsgrade *df* Formel (5.28).

Ausfallmechanismus nicht ignorierbar, so ist eine Verteilungsannahme für $\underline{R}$ gegeben $\underline{y}$ zu treffen, und der Parametervektor θ ist aus der Likelihood-Funktion $L(\mathbf{y}_{obs}, \mathbf{r} \mid \theta, \psi)$ zu schätzen. Die Maximierung dieser Funktion erweist sich, wie in den Kapiteln 6.3.2.2 und 6.3.2.3 gezeigt, bereits im univariaten Fall als kompliziert. Ein weiteres Problem betrifft die mögliche Fehlspezifikation des Verteilungstyps von $\underline{R}$ gegeben $\underline{y}$. Diese kann zu verzerrten Schätzern führen, so dass die Untersuchung verschiedener Ausfallmechanismen in der Literatur empfohlen wird.[128]

6.4 Pattern-Mixture Modelle

6.4.1 Theoretische Grundlagen

Als Pattern-Mixture Modelle bezeichnete Ansätze zur Behandlung von fehlenden Werten beruhen auf der folgenden Faktorisierung der gemeinsamen Verteilung $P(\underline{Y}, \underline{R})$, welche durch die Parametervektoren ε und ω beschrieben wird:[129]

$$P_{\omega,\varepsilon}(\underline{Y}, \underline{R}) = P_{\omega}(\underline{Y} \mid \underline{R}) P_{\varepsilon}(\underline{R}) \tag{6.50}$$

In diesem Ansatz ist die bedingte Verteilung $P_{\omega}(\underline{Y} \mid \underline{R})$ für alle Werte, welche die Zufallsvariable $\underline{R}$ annehmen kann, zu modellieren. Werden diese Ausprägungen mit $\underline{a}_0,\ldots,\underline{a}_s$ bezeichnet, so wird für jeden Wert $\underline{a}_l$ ($l = 0,\ldots, s$) ein so genanntes „Pattern“ (Muster) r durch $r = l$ definiert. Die folgende Abbildung verdeutlicht diese Festlegung an einem Beispiel.[130]

i	R_{i1}	R_{i2}	R_{i3}	R_{i4}	R_{i5}	Pattern r
1	0	0	0	0	0	0
2	0	0	1	1	0	1
3	1	0	0	0	1	2
4	0	0	1	1	0	1
5	1	1	1	1	1	3
6	1	0	0	0	1	2

Abbildung 6.3: Definition der Pattern im Pattern-Mixture Modell am Beispiel von $k = 5$ Zufallsvariablen

[128] Vgl. Toutenburg et al. (2004), S. 34.
[129] Vgl. Little (1993), S. 125.
[130] Es sei in diesem Beispiel davon ausgegangen, dass alle Ausprägungen $\underline{a}_0,\ldots,\underline{a}_s$ von $\underline{R}$ in der Stichprobe auftreten.

Insofern kann r auch als Realisation einer eindimensionalen, diskreten Zufallsvariable R aufgefasst werden und die bedingten Verteilungen $P_{\omega_r}(\underline{Y} \mid R = r)$ ($r = 0,..., s$) sind durch einen Parametervektor ω_r beschrieben.[131] Das Ziel des Pattern-Mixture Ansatzes ist die Schätzung dieser im Vektor $\omega = (\omega_0,...,\omega_s)$ zusammengefassten Parametervektoren $\omega_0,\ldots,\omega_s$, welche sich insbesondere bei der Betrachtung mehrerer Zufallsvariablen mit fehlenden Werten als kompliziert und aufwendig erweist, da in diesem Fall die Anzahl (s+1) der Pattern im Allgemeinen hoch ist.[132] Die Bezeichnung „Pattern-Mixture“ resultiert aus der Eigenschaft des Modells, dass sich die marginale Verteilung von $\underline{Y}$ aus den (s+1) verschiedenen bedingten Verteilungen gegeben die Pattern zusammensetzt.[133]

Die Mischung dieser bedingten Verteilungen wird durch die Verteilung $P_\varepsilon(\underline{R})$ bestimmt.[134] Der Parametervektor $\varepsilon = (\varepsilon_0,...,\varepsilon_s)$ gibt dabei die Wahrscheinlichkeit für jedes einzelne Pattern wieder:

$$P_{\varepsilon_r}(R = r) = \varepsilon_r$$

Die Parameter in dem Vektor ε können durch den jeweiligen Anteil von Merkmalsträgern, die zu einem bestimmten Muster gehören, an der gesamten Stichprobe geschätzt werden.[135] Vorausetzung für diese Schätzung ist, dass bei frequentistischer Sichtweise der gemeinsame, mit $\mathbf{\Omega}_{\omega,\varepsilon}$ bezeichnete Parameterraum von (ω, ε) das karthesische Kreuzprodukt des mit $\mathbf{\Omega}_\omega$ benannten Parameterraums von ω und des durch $\mathbf{\Omega}_\varepsilon$ symbolisierten Parameterraums von ε ist:[136]

$$\mathbf{\Omega}_{\omega,\varepsilon} = \mathbf{\Omega}_\omega \times \mathbf{\Omega}_\varepsilon \tag{6.51}$$

Der bayesianischen Theorie folgend ist dagegen anzunehmen, dass die Parametervektoren ω und ε a priori unabhängig sind, so dass die gemeinsame a priori Verteilung $P(\omega, \varepsilon)$ das Produkt der beiden a priori Verteilungen von ω und ε ist:

---

[131] Vgl. Schafer/Graham (2002), S. 172.
[132] Vgl. Storck et al. (2000), S. 3.
[133] Vgl. Little (1993), S. 125.
[134] Vgl. Faktorisierung der gemeinsamen Verteilung in (6.50).
[135] Vgl. Schafer/Graham (2002), S. 172.
[136] Diese Annahme ist darüber hinaus auch für die Schätzung von ω in den folgenden Kapiteln notwendig.

$$P(\omega, \varepsilon) = P(\omega)P(\varepsilon) \tag{6.52}$$

Für die weiteren Ausführungen in diesem Kapitel sei angenommen, dass diese Bedingungen erfüllt sind.[137] Ein Vorteil dieses Modellansatzes gegenüber den Selection Modellen ist, dass die Momente von $\underline{Y}$ in Abhängigkeit von den nicht identifizierbaren Parametern des Vektors ω geschätzt werden können. Das folgende Beispiel veranschaulicht diesen Aspekt.

Beispiel 6.7:[138]

Es wird eine binäre Zufallsvariable $\underline{Y}$ betrachtet, deren Erwartungswert aus einer Stichprobe vom Umfang n mit (n-q) fehlenden Werten zu schätzen ist. Dies kann über die Pattern-Mixture Modellierung

$$P_{\omega,\varepsilon}(\underline{Y}, \underline{R}) = P_{\omega}(\underline{Y} \mid \underline{R})P_{\varepsilon}(\underline{R}) \qquad \omega = (\omega_0, \omega_1), \varepsilon = (\varepsilon_0, \varepsilon_1)$$

aus der Randverteilung von $\underline{Y}$

$$P_{\omega,\varepsilon}(\underline{Y}) = \sum_{\underline{r}=0}^{1} P_{\omega,\varepsilon}(\underline{Y}, \underline{R} = \underline{r})$$

erfolgen:[139]

$$\begin{aligned} E(\underline{Y}) &= P_{\omega,\varepsilon}(\underline{Y} = 1) \\ &= P_{\omega_0}(\underline{Y} = 1 \mid \underline{R} = 0)P_{\varepsilon_0}(\underline{R} = 0) + P_{\omega_1}(\underline{Y} = 1 \mid R = 1)P_{\varepsilon_1}(R = 1) \\ &= P_{\omega_0}(\underline{Y} = 1 \mid \underline{R} = 0)P_{\varepsilon_0}(\underline{R} = 0) + P_{\omega_1}(\underline{Y} = 1 \mid \underline{R} = 1)\left[1 - P_{\varepsilon_0}(\underline{R} = 0)\right] \\ &= \omega_0\varepsilon_0 + \omega_1(1 - \varepsilon_0) \end{aligned} \tag{6.53}$$

Der Parameter ε_0 der Verteilung $P_{\varepsilon}(\underline{R})$ ($\varepsilon = (\varepsilon_0, \varepsilon_1)$, $\varepsilon_1 = 1\text{-}\varepsilon_0$) kann aus dem Anteil der vollständigen Untersuchungseinheiten am Stichprobenumfang durch

$$\hat{\varepsilon}_0 = \frac{q}{n}$$

[137] Innerhalb der Selection Modelle ist von den gleichen Bedingungen für die Parameterräume bzw. a priori Verteilungen von θ und ψ ausgegangen worden.
[138] Vgl. Toutenburg et al. (2004), S. 31f.
[139] Vgl. Toutenburg et al. (2004), S. 31.

geschätzt werden. Der Schätzer für ω_0 ist ebenfalls aus den beobachteten Daten durch

$$\hat{\omega}_0 = \frac{1}{q}\sum_{i=1}^{q} y_i$$

bestimmbar. Im Pattern $\underline{r} = 1$ sind hingegen keine Werte beobachtet worden, so dass eine Schätzung des Parameters ω_1 aus den Daten der Stichprobe nicht möglich ist. Dennoch kann durch Variierung von $\hat{\omega}_1$ $(0 \le \hat{\omega}_1 \le 1)$ zumindest die Sensitivität der Schätzung für den Erwartungswert $E(\underline{Y})$ beurteilt werden. Darüber hinaus erhält man durch Einsetzen der Extremwerte $\omega_1 = 0$ und $\omega_1 = 1$ in (6.53) die folgende Beschränkung für den Erwartungswert von $\underline{Y}$:[140]

$$\omega_0\varepsilon_0 \le \mathrm{E}(\underline{Y}) \le 1-\left[(1-\omega_0)\varepsilon_0\right] \tag{6.54}$$

Durch die Pattern-Mixture Modellierung ist es somit ebenfalls möglich, einen Wertebereich für den Schätzer eines Moments zu bestimmen. Aus (6.54) wird ersichtlich, dass dies insbesondere in den Fällen sinnvoll ist, in denen der Anteil vollständiger Datensätze $\hat{\varepsilon}_0 = \frac{q}{n}$ hoch ist.

Im Vergleich zu diesem Beispiel erweist sich die Pattern-Mixture Modellierung und die anschließende Schätzung der Momente von $\underline{Y}$ als deutlich komplizierter, falls mehrere Zufallsvariablen $Y_1,...,Y_k$ $(k > 1)$ in dem Modell betrachtet werden. Die folgenden Ausführungen in Kapitel 6.4.2 zeigen, wie in diesem Fall auch bei Vorliegen eines nicht ignorierbaren Ausfallmechanismus die Momente unverzerrt geschätzt werden können, wenn entsprechende Annahmen erfüllt sind. Diese Schätzung erfolgt aus den Parameterschätzern $\hat{\omega}$ und $\hat{\varepsilon}$ der gemeinsamen Verteilung $P_{\omega,\varepsilon}(\mathbf{Y},\mathbf{R})$, für die aufgrund der iid-Eigenschaft der Stichprobe allgemein im Pattern-Mixture Modell gilt:

$$P_{\omega,\varepsilon}(\mathbf{Y},\mathbf{R}) = \prod_{i=1}^{n} P_{\omega,\varepsilon}(\underline{Y}_i,\underline{R}_i) = \prod_{i=1}^{n} P_{\omega}(\underline{Y}_i \mid \underline{R}_i) P_{\varepsilon}(\underline{R}_i) = P_{\omega}(\mathbf{Y} \mid \mathbf{R}) P_{\varepsilon}(\mathbf{R})$$

[140] Vgl. Toutenburg et al. (2004), S. 32.

In Kapitel 6.4.3 wird insbesondere auf die Problematik von Unit Nonresponse eingegangen, bei der alle Realisationen $y_{i1},...,y_{ik}$ in einem Datensatz i ($1 \le i \le n$) nicht beobachtet wurden. Innerhalb von Pattern-Mixture Modellen wird Unit Nonresponse als ein konkretes Pattern definiert, während die weiteren Muster das Fehlen einzelner Werte in einem Datensatz (Item Nonresponse) indizieren. Diese strikte Trennung ermöglicht es, verschiedene Ausfallmechanismen für Unit Nonresponse und Item Nonresponse zu berücksichtigen. An einem abschließenden Beispiel wird gezeigt, wie dieser methodische Vorteil der Pattern-Mixture Modelle gegenüber den Selection Modellen praktisch umgesetzt werden kann.

6.4.2 Bivariates Pattern-Mixture Modell

6.4.2.1 Vorbetrachtung

Im Beispiel 6.7 wurde eine eindimensionale diskrete Zufallsvariable $\underline{Y}$ betrachtet, deren Erwartungswert E($\underline{Y}$) aus den unvollständig beobachteten Daten geschätzt werden soll. Im Folgenden sei eine stetige bivariate Zufallsvariable $\underline{Y} = (Y_1, Y_2)$ der Ausgangspunkt, deren Momente aus den Werten einer einfachen Stichprobe vom Umfang n zu schätzen ist. In dieser Stichprobe seien die Werte $y_{11},...,y_{n1}$ der Zufallsvariable Y_1 vollständig beobachtet worden, während von den Zufallsvariablen $Y_{12},...,Y_{n2}$ lediglich die Werte $y_{12},...,y_{q2}$ bekannt sind ($q < n$). Little (1994) betrachtet für diesen Fall ein Pattern-Mixture Modell mit den folgenden Verteilungsannahmen:[141]

$$(1)\quad (Y_1, Y_2 \mid R_2 = r) \sim N\left(\mu^{(r)}, \Sigma^{(r)}\right) \qquad r = 0,1$$

$$(2)\quad R_2 \sim B(1, \varepsilon_1) \tag{6.55}$$

Im Beispiel des vorliegenden univariaten Datenausfalls werden die Merkmalsträger in zwei Pattern eingeteilt: Das erste Muster ($r = 0$) umfasst die Merkmalsträger, die vollständig beobachtet wurden, und das zweite Pattern ($r = 1$) beinhaltet die Datensätze, bei denen y_{i2} nicht beobachtet wurde. Es wird angenommen, dass die Zufallsvariable (Y_1, Y_2) bivariat normalverteilt *innerhalb dieser beiden Pattern* ist, jedoch

[141] Vgl. Little (1994), S. 474.

mit unterschiedlichen Parametervektoren

$$\mu^{(r)} = \left(\mu_1^{(r)}, \mu_2^{(r)}\right),\ \Sigma^{(r)} = \begin{pmatrix} \sigma_{11}^{(r)} & \sigma_{12}^{(r)} \\ \sigma_{21}^{(r)} & \sigma_{22}^{(r)} \end{pmatrix}$$

für r = 0,1. Dementsprechend seien die Zufallsvariablen $Y_j^{(r)}$ (j = 1, 2 ; r = 0, 1) in der folgenden Weise definiert:

$$Y_j^{(r)} := \quad Y_j \mid R_2 = r \tag{6.56}$$

Die Parameter $\mu^{(r)}$ und $\Sigma^{(r)}$ sind im Pattern-Mixture Ansatz die Elemente des Vektors ω_r (r = 0,1):

$$\omega_r = \left(\mu^{(r)}, \Sigma^{(r)}\right)$$

Die im Vektor $\omega = (\omega_0, \omega_1)$ zusammengefassten Parametervektoren beschreiben im Pattern-Mixture Modell die bedingte Verteilung $P_\omega(Y_1, Y_2 \mid R_2)$, während ε der Parameter der Verteilung $P_\varepsilon(R_2)$ ist ($\varepsilon = (\varepsilon_0, \varepsilon_1)$, $\varepsilon_0 = 1\text{-}\varepsilon_1$). Für die gemeinsame Verteilung $P_{\omega,\varepsilon}(Y_1, Y_2, R_2)$ gilt in diesem Modell:[142]

$$P_{\omega,\varepsilon}(Y_1, Y_2, R_2) = P_\omega(Y_1, Y_2 \mid R_2) P_\varepsilon(R_2)$$

Ausgehend von dieser Faktorisierung und unter Berücksichtigung der iid-Eigenschaft der Stichprobe gilt für $P_{\omega,\varepsilon}(\mathbf{y}, \mathbf{r})$:

$$\begin{aligned} P_{\omega,\varepsilon}(\mathbf{y}, \mathbf{r}) &= \prod_{i=1}^{n} P_{\omega_r}(Y_{i1} = y_{i1}, Y_{i2} = y_{i2} \mid R_{i2} = r) P_{\varepsilon_r}(R_{i2} = r) \\ &= \prod_{i=1}^{q} P_{\omega_0}(Y_{i1} = y_{i1}, Y_{i2} = y_{i2} \mid R_{i2} = 0) P_{\varepsilon_0}(R_{i2} = 0) \\ &\quad \cdot \prod_{i=q+1}^{n} P_{\omega_1}(Y_{i1} = y_{i1}, Y_{i2} = y_{i2} \mid R_{i2} = 1) P_{\varepsilon_1}(R_{i2} = 1) \\ &= (1-\varepsilon_1)^q \prod_{i=1}^{q} P_{\omega_0}(Y_{i1} = y_{i1}, Y_{i2} = y_{i2} \mid R_{i2} = 0) \\ &\quad \cdot (\varepsilon_1)^{n-q} \prod_{i=q+1}^{n} P_{\omega_1}(Y_{i1} = y_{i1}, Y_{i2} = y_{i2} \mid R_{i2} = 1) \end{aligned} \tag{6.57}$$

[142] Vgl. allgemeines Modell in (6.50).

Durch Integration von $P_{\omega,\varepsilon}(\mathbf{y},\mathbf{r})$ über die fehlenden Werte $\mathbf{y}_{mis} = (y_{(q+1)2},...,y_{n2})$ erhält man $P_{\omega,\varepsilon}(\mathbf{y}_{obs},\mathbf{r})$:[143]

$$\begin{aligned} P_{\omega,\varepsilon}(\mathbf{y}_{obs},\mathbf{r}) &= (1-\varepsilon_1)^q \prod_{i=1}^{q} P_{\omega_0}(Y_{i1}=y_{i1}, Y_{i2}=y_{i2} \mid R_{i2}=0) \\ &\quad \cdot (\varepsilon_1)^{n-q} \prod_{i=q+1}^{n} \int P_{\omega_1}(Y_{i1}=y_{i1}, Y_{i2}=y_{i2} \mid R_{i2}=1)\, dy_{i2} \\ &= (1-\varepsilon_1)^q \prod_{i=1}^{q} P_{\omega_0}(Y_{i1}=y_{i1}, Y_{i2}=y_{i2} \mid R_{i2}=0) \\ &\quad \cdot (\varepsilon_1)^{n-q} \prod_{i=q+1}^{n} P_{\mu_1^{(1)},\sigma_{11}^{(1)}}(Y_{i1}=y_{i1} \mid R_{i2}=1) \end{aligned} \tag{6.58}$$

Die bedingte Verteilung $P_{\mu_1^{(1)},\sigma_{11}^{(1)}}(Y_1 \mid R_2 = 1)$ ist eine univariate Normalverteilung mit den Parametern $\mu_1^{(1)}$ und $\sigma_{11}^{(1)}$.[144] Anhand von (6.58) wird ersichtlich, dass die entsprechende Likelihood-Funktion

$$\begin{aligned} L(\mathbf{y}_{obs},\mathbf{r} \mid \omega,\varepsilon) &= (1-\varepsilon_1)^q \prod_{i=1}^{q} P_{\omega_0}(Y_{i1}=y_{i1}, Y_{i2}=y_{i2} \mid R_{i2}=0) \\ &\quad \cdot (\varepsilon_1)^{n-q} \prod_{i=q+1}^{n} P_{\mu_1^{(1)},\sigma_{11}^{(1)}}(Y_{i1}=y_{i1} \mid R_{i2}=1) \end{aligned}$$

nur von den Parametern $\omega_0 = \left(\mu_1^{(0)}, \mu_2^{(0)}, \sigma_{11}^{(0)}, \sigma_{22}^{(0)}, \sigma_{12}^{(0)}\right)$, $\mu_1^{(1)}$, $\sigma_{11}^{(1)}$ und ε_1 abhängig ist. Diese Parameter können durch die Maximum-Likelihood-Methode aus den beobachteten Daten geschätzt werden. Für die Bestimmung des Schätzers $\hat{\varepsilon}_1$ ist dabei die partielle Ableitung der Loglikelihood-Funktion nach ε_1 zu bilden:

$$\frac{\partial \ln L(\mathbf{y}_{obs},\mathbf{r} \mid \omega,\varepsilon)}{\partial \varepsilon_1} = \frac{-q}{1-\varepsilon_1} + \frac{n-q}{\varepsilon_1}$$

Durch Nullsetzen dieser partiellen Ableitung ist der Maximum-Likelihood-Schätzer $\hat{\varepsilon}_1$ zu bestimmen:

$$\begin{aligned} &\frac{-q}{1-\hat{\varepsilon}_1} + \frac{n-q}{\hat{\varepsilon}_1} = 0 \\ &\Rightarrow \hat{\varepsilon}_1 = \frac{n-q}{n} \end{aligned} \tag{6.59}$$

[143] Vgl. Little/Rubin (2002), S. 332.
[144] Vgl. Hartung (1999), S. 120.

In analoger Weise erhält man die ML-Schätzer für den Parametervektor ω_0 sowie die Parameter $\mu_1^{(1)}$ und $\sigma_{11}^{(1)}$:[145]

$$\hat{\mu}_j^{(0)} = \frac{1}{q}\sum_{i=1}^{q} y_{ij} = \bar{y}_j^{(0)} \qquad \forall j = 1,2 \tag{6.60}$$

$$\hat{\mu}_1^{(1)} = \frac{1}{n-q}\sum_{i=q+1}^{n} y_{i1} = \bar{y}_1^{(1)} \tag{6.61}$$

$$\hat{\sigma}_{jl}^{(0)} = \frac{1}{q}\sum_{i=1}^{q} (y_{ij} - \bar{y}_j^{(0)})(y_{il} - \bar{y}_l^{(0)}) \qquad \forall j = 1,2; l = 1,2 \tag{6.62}$$

$$\hat{\sigma}_{11}^{(1)} = \frac{1}{n-q}\sum_{i=q+1}^{n} (y_{i1} - \bar{y}_1^{(1)})^2 \tag{6.63}$$

Drei Parameter von $P_{\omega,\varepsilon}(\mathbf{y},\mathbf{r})$ aus (6.57) können nicht ohne weitere Annahmen geschätzt werden. Dies sind die Parameter $\mu_2^{(1)}, \sigma_{22}^{(1)}$ und $\sigma_{12}^{(1)}$, welche (zusammen mit den schätzbaren Parametern $\mu_1^{(1)}$ und $\sigma_{11}^{(1)}$) die bedingte Dichte

$$P_{\omega_1}(Y_2 = y_2, Y_1 = y_1 \mid R_2 = 1)$$

bestimmen. Im Weiteren wird gezeigt, wie unter verschiedenen Annahmen eine Schätzung dieser Parameter erfolgen kann. Dabei wird die Vorgehensweise zunächst anhand der MAR-Annahme verdeutlicht, während anschließend ein spezieller, nicht ignorierbarer Ausfallmechanismus untersucht wird.

6.4.2.2 Bivariates Pattern-Mixture Modell unter der MAR-Annahme

Die MAR-Annahme impliziert die bedingte Unabhängigkeitsbeziehung

$R_2 \perp\!\!\!\perp Y_2 \mid Y_1$,

so dass die bedingten Dichten $P_{\omega_{0,2\cdot1}}(Y_2 \mid Y_1 = y_1, R_2 = 0)$ und $P_{\omega_{1,2\cdot1}}(Y_2 \mid Y_1 = y_1, R_2 = 1)$ bei der Erfüllung dieser Annahme identisch sind:

$$P_{\omega_{0,2\cdot1}}(Y_2 \mid Y_1 = y_1, R_2 = 0) = P_{\omega_{1,2\cdot1}}(Y_2 \mid Y_1 = y_1, R_2 = 1)$$
$$\Leftrightarrow P_{\omega_{0,2\cdot1}}(Y_2^{(0)} \mid Y_1^{(0)} = y_1^{(0)}) = P_{\omega_{1,2\cdot1}}(Y_2^{(1)} \mid Y_1^{(1)} = y_1^{(1)}) \tag{6.64}$$

---

[145] Vgl. Little (1994), S. 475.

Die Parametervektoren $\omega_{r,2\cdot1} = \left(a^{(r)}, b^{(r)}, \sigma_{22\cdot1}^{(r)}\right)$ (r = 0, 1) dieser bedingten Dichten umfassen dabei die Parameter der einfachen linearen Regressionsmodelle

$$Y_2^{(r)} = a^{(r)} + b^{(r)} Y_1^{(r)} + U^{(r)} \qquad (r = 0,1)$$

mit den Zufallsvariablen $U^{(r)}$, welche den Erwartungswert $E\left(U^{(r)}\right) = 0$ und die Varianz $\mathrm{Var}\left(U^{(r)}\right) = \sigma_{22\cdot1}^{(r)}$ besitzen.[146]

Basierend auf der Methode der Kleinsten Quadrate erhält man die folgenden Schätzer für die Parameter $b^{(r)}$ (r = 0, 1) der Regressionsmodelle:

$$\hat{b}^{(r)} = \frac{\hat{\sigma}_{12}^{(r)}}{\hat{\sigma}_{11}^{(r)}} \qquad (r = 0,1) \tag{6.65}$$

Weiterhin kann die Varianz der Residuen

$$\begin{aligned} \mathrm{Var}\left(U^{(r)}\right) &= \mathrm{Var}\left(Y_2^{(r)}\right) - \mathrm{Var}\left(a^{(r)} + b^{(r)} Y_1^{(r)}\right) \\ &= \mathrm{Var}\left(Y_2^{(r)}\right) - \left(b^{(r)}\right)^2 \mathrm{Var}\left(Y_1^{(r)}\right) \qquad (r = 0,1) \end{aligned}$$

durch

$$\hat{\sigma}_{22\cdot1}^{(r)} = \hat{\sigma}_{22}^{(r)} - \left(\frac{\hat{\sigma}_{12}^{(r)}}{\hat{\sigma}_{11}^{(r)}}\right)^2 \hat{\sigma}_{11}^{(r)} = \hat{\sigma}_{22}^{(r)} - \frac{\left(\hat{\sigma}_{12}^{(r)}\right)^2}{\hat{\sigma}_{11}^{(r)}} \qquad (r = 0,1) \tag{6.66}$$

geschätzt werden, und aus

$$\mu_2^{(r)} = E(Y_2^{(r)}) = E(a^{(r)} + b^{(r)} Y_1^{(r)} + U^{(r)}) = a^{(r)} + b^{(r)} \mu_1^{(r)} \qquad (r = 0,1)$$

folgt für den Schätzer $\hat{a}^{(r)}$:

$$\hat{a}^{(r)} = \hat{\mu}_2^{(r)} - \hat{b}^{(r)} \hat{\mu}_1^{(r)} \qquad (r = 0,1) \tag{6.67}$$

Da in diesem Kapitel die Gültigkeit der MAR-Annahme vorausgesetzt wird, gilt für die Parametervektoren $\omega_{0,2\cdot1} = \left(a^{(0)}, b^{(0)}, \sigma_{22\cdot1}^{(0)}\right)$ und $\omega_{1,2\cdot1} = \left(a^{(1)}, b^{(1)}, \sigma_{22\cdot1}^{(1)}\right)$ aufgrund von (6.64):

$$\omega_{0,2\cdot1} = \omega_{1,2\cdot1} \tag{6.68}$$

---

[146] Vgl. Anderson (1957), S. 200.

Durch diese Identität der Vektoren ist es ausreichend, die Parameter $b^{(0)}$, $\sigma_{22\cdot1}^{(0)}$ und $a^{(0)}$ mittels (6.65), (6.66) und (6.67) aus den beobachteten Daten zu schätzen, denn für die Parameterschätzer im Pattern $r = 1$ gilt:

$$\hat{a}^{(1)} = \hat{a}^{(0)} \quad , \qquad \hat{b}^{(1)} = \hat{b}^{(0)} \quad , \qquad \hat{\sigma}_{22\cdot1}^{(1)} = \hat{\sigma}_{22\cdot1}^{(0)} \tag{6.69}$$

Ausgehend von den Schätzern $\hat{\omega}_{r,2\cdot1} = \left(\hat{a}^{(r)}, \hat{b}^{(r)}, \hat{\sigma}_{22\cdot1}^{(r)}\right)$ ($r = 0, 1$) können auch die Parameter $\mu_2^{(1)}, \sigma_{22}^{(1)}$ und $\sigma_{12}^{(1)}$ der gemeinsamen Dichtefunktion

$$P_{\omega_1}(Y_1^{(1)}, Y_2^{(1)}) \qquad \left(\omega_1 = \left(\mu_1^{(1)}, \mu_2^{(1)}, \sigma_{11}^{(1)}, \sigma_{22}^{(1)}, \sigma_{12}^{(1)}\right)\right)$$

geschätzt werden:[147]

$$\begin{aligned}
\hat{\mu}_2^{(1)} &= \hat{a}^{(1)} + \hat{b}^{(1)}\hat{\mu}_1^{(1)} \\
&= \hat{a}^{(0)} + \hat{b}^{(0)}\hat{\mu}_1^{(1)} \\
&= \left(\hat{\mu}_2^{(0)} - \hat{b}^{(0)}\hat{\mu}_1^{(0)}\right) + \hat{b}^{(0)}\hat{\mu}_1^{(1)} \\
&= \bar{y}_2^{(0)} + \hat{b}^{(0)}\left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right)
\end{aligned} \tag{6.70}$$

$$\begin{aligned}
\hat{\sigma}_{22}^{(1)} &= \hat{\sigma}_{22\cdot1}^{(1)} + \left(\hat{b}^{(1)}\right)^2 \hat{\sigma}_{11}^{(1)} \\
&= \hat{\sigma}_{22\cdot1}^{(0)} + \left(\hat{b}^{(0)}\right)^2 \hat{\sigma}_{11}^{(1)} \\
&= \hat{\sigma}_{22}^{(0)} - \left(\hat{b}^{(0)}\right)^2 \hat{\sigma}_{11}^{(0)} + \left(\hat{b}^{(0)}\right)^2 \hat{\sigma}_{11}^{(1)} \\
&= \hat{\sigma}_{22}^{(0)} + \left(\hat{b}^{(0)}\right)^2 \left(\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11}^{(0)}\right)
\end{aligned} \tag{6.71}$$

$$\hat{\sigma}_{12}^{(1)} = \hat{b}^{(1)}\hat{\sigma}_{11}^{(1)} = \hat{b}^{(0)}\hat{\sigma}_{11}^{(1)} \tag{6.72}$$

Darüber hinaus ist eine unverzerrte Schätzung der Momente von Y_2 sowie der Kovarianz von Y_1 und Y_2 (σ_{12}) unter der MAR-Annahme möglich. Da die Randdichte der Zufallsvariable Y_2 eine Mischung der beiden Dichtefunktionen $P_{\mu_2^{(0)},\sigma_{22}^{(0)}}(Y_2^{(0)})$ und $P_{\mu_2^{(1)},\sigma_{22}^{(1)}}(Y_2^{(1)})$ ist, besitzt Y_2 den folgenden Erwartungswert:[148]

$$E(Y_2) = \mu_2 = (1 - \varepsilon_1)\mu_2^{(0)} + \varepsilon_1\mu_2^{(1)} \tag{6.73}$$

---

[147] Die Parameter $\mu_1^{(1)}$ und $\sigma_{11}^{(1)}$ können ohne jegliche Annahmen aus den Daten geschätzt werden (vgl. Schätzer in (6.61) und (6.63)).

[148] Vgl. Little/Rubin (2002), S. 332.

Durch Einsetzen der in (6.59) und (6.61) hergeleiteten Schätzer und unter Berücksichtigung von (6.70) kann $\hat{\mu}_2$ bestimmt werden:[149]

$$\begin{aligned}\hat{\mu}_2 &= (1-\hat{\varepsilon}_1)\hat{\mu}_2^{(0)} + \hat{\varepsilon}_1\,\hat{\mu}_2^{(1)} \\ &= \frac{q}{n}\bar{y}_2^{(0)} + \frac{n-q}{n}\left(\bar{y}_2^{(0)} + \hat{b}^{(0)}\left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right)\right) \\ &= \bar{y}_2^{(0)} + \frac{n-q}{n}\hat{b}^{(0)}\left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right) \\ &= \bar{y}_2^{(0)} + \hat{b}^{(0)}\left(\bar{y}_1 - \frac{q}{n}\bar{y}_1^{(0)} - \frac{n-q}{n}\bar{y}_1^{(0)}\right) \qquad ;\ \bar{y}_1 = \frac{1}{n}\sum_{i=1}^{n} y_{i1} \\ &= \bar{y}_2^{(0)} + \hat{b}^{(0)}\left(\bar{y}_1 - \bar{y}_1^{(0)}\right) \end{aligned} \tag{6.74}$$

Die Kovarianz von Y_1 und Y_2 ist durch

$$\hat{\sigma}_{12} = \hat{b}^{(0)}\hat{\sigma}_{11} \qquad ;\ \hat{\sigma}_{11} = \frac{1}{n}\sum_{i=1}^{n}(y_{i1} - \bar{y}_1)^2 \tag{6.75}$$

zu schätzen, während für den Schätzer der Varianz von Y_2 gilt:[150]

$$\begin{aligned}\hat{\sigma}_{22} &= \hat{\sigma}_{22\cdot 1}^{(0)} + \left(\hat{b}^{(0)}\right)^2\hat{\sigma}_{11} \\ &= \hat{\sigma}_{22}^{(0)} - \left(\hat{b}^{(0)}\right)^2\hat{\sigma}_{11}^{(0)} + \left(\hat{b}^{(0)}\right)^2\hat{\sigma}_{11} \\ &= \hat{\sigma}_{22}^{(0)} + \left(\hat{b}^{(0)}\right)^2\left(\hat{\sigma}_{11} - \hat{\sigma}_{11}^{(0)}\right) \end{aligned} \tag{6.76}$$

Die Schätzungen in (6.74)-(6.76) beruhen auf der Annahme, dass der Ausfallmechanismus vom Typ MAR ist. Diese Annahme wurde im Pattern-Mixture Modell in der Weise berücksichtigt, dass basierend auf dem Zusammenhang

$$\omega_{0,2\cdot 1} = \omega_{1,2\cdot 1}$$

eine Schätzung der – ohne Annahmen nicht identifizierbaren – Parameter $\mu_2^{(1)}, \sigma_{22}^{(1)}$ und $\sigma_{12}^{(1)}$ erfolgt. Solche Annahmen, die eine Schätzung der ursprünglich nicht identifizierbaren Parameter in den Pattern erlauben, werden von Little (1993) allgemein als „Identifying Restrictions“ bezeichnet.[151] Im Folgenden wird anhand einer speziellen Restriktion gezeigt, wie auch bei nicht ignorierbarem Ausfallmechanismus eine Schätzung aller Parameter des Pattern-Mixture Modells erfolgen kann.

---

[149] Vgl. Little (1994), S. 475.
[150] Vgl. Little/Rubin (2002), S. 137.
[151] Vgl. Little (1993), S. 127.

6.4.2.3 Bivariates Pattern-Mixture Modell unter MNAR

Hängt der Datenausfall – gegeben die Zufallsvariable Y_1 – nur von Y_2 selbst ab, so gilt die bedingte Unabhängigkeitsbeziehung

$$R_2 \perp\!\!\!\perp Y_1 \mid Y_2 \tag{6.77}$$

und der Ausfallmechanismus ist vom Typ MNAR. In diesem Fall sind die bedingten Dichten $P_{\omega_{0,1\cdot2}}(Y_1 \mid Y_2 = y_2, R_2 = 0)$ und $P_{\omega_{1,1\cdot2}}(Y_1 \mid Y_2 = y_2, R_2 = 1)$ identisch:

$$P_{\omega_{0,1\cdot2}}(Y_1 \mid Y_2 = y_2, R_2 = 0) = P_{\omega_{1,1\cdot2}}(Y_1 \mid Y_2 = y_2, R_2 = 1)$$

$$\Leftrightarrow P_{\omega_{0,1\cdot2}}(Y_1^{(0)} \mid Y_2^{(0)} = y_2^{(0)}) = P_{\omega_{1,1\cdot2}}(Y_1^{(1)} \mid Y_2^{(1)} = y_2^{(1)})$$

Die Parametervektoren $\omega_{r,1\cdot2} = \left(c^{(r)}, d^{(r)}, \sigma_{11\cdot2}^{(r)}\right)$ (r = 0, 1) dieser bedingten Dichten beziehen sich auf die Parameter der einfachen linearen Regressionsmodelle

$$Y_1^{(r)} = c^{(r)} + d^{(r)} Y_2^{(r)} + V^{(r)} \qquad (r = 0,1)\,.$$

Die Zufallsvariablen $V^{(r)}$ (r = 0, 1) besitzen den Erwartungswert $E\left(V^{(r)}\right) = 0$ und die Varianz $\mathrm{Var}\left(V^{(r)}\right) = \sigma_{11\cdot2}^{(r)}$.

In einem einfachen linearen Regressionsmodell ergeben sich die folgenden Schätzer für die Parameter $d^{(r)}$, $c^{(r)}$ und die Varianz $\sigma_{11\cdot2}^{(r)}$:

$$\hat{d}^{(r)} = \frac{\hat{\sigma}_{12}^{(r)}}{\hat{\sigma}_{22}^{(r)}}$$

$$\hat{c}^{(r)} = \hat{\mu}_1^{(r)} - \hat{d}^{(r)} \hat{\mu}_2^{(r)}$$

$$\hat{\sigma}_{11\cdot2}^{(r)} = \hat{\sigma}_{11}^{(r)} - \frac{\left(\hat{\sigma}_{12}^{(r)}\right)^2}{\hat{\sigma}_{22}^{(r)}} \qquad (r = 0,1) \tag{6.78}$$

Ist die Annahme in (6.77) erfüllt, so sind die Parametervektoren $\omega_{0,1\cdot2} = \left(c^{(0)}, d^{(0)}, \sigma_{11\cdot2}^{(0)}\right)$ und $\omega_{1,1\cdot2} = \left(c^{(1)}, d^{(1)}, \sigma_{11\cdot2}^{(1)}\right)$ identisch:

$$\omega_{0,1\cdot2} = \omega_{1,1\cdot2} \tag{6.79}$$

Unter dieser Restriktion können die Parameter $\mu_2^{(1)}, \sigma_{22}^{(1)}$ und $\sigma_{12}^{(1)}$ der gemeinsamen Dichte $P_{\omega_1}(Y_1^{(1)}, Y_2^{(1)})$ aus den beobachteten Daten geschätzt werden:[152, 153]

$$\hat{\mu}_2^{(1)} = \frac{\hat{\mu}_1^{(1)} - \hat{c}^{(1)}}{\hat{d}^{(1)}} = \frac{\hat{\mu}_1^{(1)} - \hat{c}^{(0)}}{\hat{d}^{(0)}} = \frac{\bar{y}_1^{(1)} - \left(\bar{y}_1^{(0)} - \hat{d}^{(0)}\bar{y}_2^{(0)}\right)}{\hat{d}^{(0)}} = \bar{y}_2^{(0)} + \frac{\bar{y}_1^{(1)} - \bar{y}_1^{(0)}}{\hat{d}^{(0)}} \tag{6.80}$$

$$\begin{aligned}\hat{\sigma}_{22}^{(1)} &= \frac{\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11\cdot 2}^{(1)}}{\left(\hat{d}^{(1)}\right)^2} = \frac{\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11\cdot 2}^{(0)}}{\left(\hat{d}^{(0)}\right)^2} = \frac{\hat{\sigma}_{11}^{(1)} - \left(\hat{\sigma}_{11}^{(0)} - \left(\hat{d}^{(0)}\right)^2 \hat{\sigma}_{22}^{(0)}\right)}{\left(\hat{d}^{(0)}\right)^2} \\ &= \hat{\sigma}_{22}^{(0)} + \frac{\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11}^{(0)}}{\left(\hat{d}^{(0)}\right)^2} \qquad (\hat{\sigma}_{11}^{(1)} \geq \hat{\sigma}_{11\cdot 2}^{(0)})\end{aligned} \tag{6.81}$$

$$\hat{\sigma}_{12}^{(1)} = \hat{d}^{(1)}\hat{\sigma}_{22}^{(1)} = \hat{d}^{(0)}\left(\hat{\sigma}_{22}^{(0)} + \frac{\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11}^{(0)}}{\left(\hat{d}^{(0)}\right)^2}\right) = \hat{\sigma}_{12}^{(0)} + \frac{\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11}^{(0)}}{\hat{d}^{(0)}} \tag{6.82}$$

Die Schätzung der Varianz $\sigma_{22}^{(1)}$ in Gleichung (6.81) führt nur unter der Bedingung $\hat{\sigma}_{11}^{(1)} \geq \hat{\sigma}_{11\cdot 2}^{(0)}$ zu zulässigen Lösungen $\hat{\sigma}_{22}^{(1)} \geq 0$. Im anderen Fall $(\hat{\sigma}_{11}^{(1)} < \hat{\sigma}_{11\cdot 2}^{(0)})$ sind die Schätzer für die Varianz und Kovarianz mit $\hat{\sigma}_{22}^{(1)} = 0$ und $\hat{\sigma}_{12}^{(1)} = 0$ festzulegen.[154]

Neben den Parametern $\mu_2^{(1)}, \sigma_{22}^{(1)}$ und $\sigma_{12}^{(1)}$ können auch die Momente der Zufallsvariable Y_2 sowie die Kovarianz von Y_1 und Y_2 unverzerrt geschätzt werden. Unter Berücksichtigung der Restriktion (6.79) und der Schätzer in (6.78) gilt:[155]

$$\begin{aligned}\hat{\mu}_2 &= \left(1 - \hat{\varepsilon}_1\right)\hat{\mu}_2^{(0)} + \hat{\varepsilon}_1\hat{\mu}_2^{(1)} \\ &= \frac{q}{n}\bar{y}_2^{(0)} + \frac{n-q}{n}\left(\bar{y}_2^{(0)} + \frac{\bar{y}_1^{(1)} - \bar{y}_1^{(0)}}{\hat{d}^{(0)}}\right) \\ &= \bar{y}_2^{(0)} + \frac{\left(\bar{y}_1 - \frac{q}{n}\bar{y}_1^{(0)}\right) - \frac{n-q}{n}\bar{y}_1^{(0)}}{\hat{d}^{(0)}} \\ &= \bar{y}_2^{(0)} + \frac{\bar{y}_1 - \bar{y}_1^{(0)}}{\hat{d}^{(0)}}\end{aligned} \tag{6.83}$$

---

[152] Vgl. Little (1994), S. 475.

[153] Vgl. Herleitung von $\hat{\omega}_0$, $\hat{\mu}_1^{(1)}$ und $\hat{\sigma}_{11}^{(1)}$ in (6.60)-(6.63).

[154] Vgl. Little/Rubin (2002), S. 333.

[155] Vgl. Little (1994), S. 475.

$$\hat{\sigma}_{22} = \frac{\hat{\sigma}_{11} - \hat{\sigma}_{11\cdot 2}^{(0)}}{\left(\hat{d}^{(0)}\right)^2} = \frac{\hat{\sigma}_{11} - \left(\hat{\sigma}_{11}^{(0)} - \left(\hat{d}^{(0)}\right)^2 \hat{\sigma}_{22}^{(0)}\right)}{\left(\hat{d}^{(0)}\right)^2} = \hat{\sigma}_{22}^{(0)} + \frac{\hat{\sigma}_{11} - \hat{\sigma}_{11}^{(0)}}{\left(\hat{d}^{(0)}\right)^2} \tag{6.84}$$

$$\hat{\sigma}_{12} = \hat{d}^{(0)}\hat{\sigma}_{22} = \hat{d}^{(0)}\left(\hat{\sigma}_{22}^{(0)} + \frac{\hat{\sigma}_{11} - \hat{\sigma}_{11}^{(0)}}{\left(\hat{d}^{(0)}\right)^2}\right) = \hat{\sigma}_{12}^{(0)} + \frac{\hat{\sigma}_{11} - \hat{\sigma}_{11}^{(0)}}{\hat{d}^{(0)}} \tag{6.85}$$

Beispiel 6.8:

Die Zufallsvariable $(Y_1^{(r)}, Y_2^{(r)})$ sei für $r = 0,1$ zweidimensional normalverteilt mit den unbekannten Parametern $\mu_1^{(r)}, \sigma_{11}^{(r)}, \mu_2^{(r)}, \sigma_{22}^{(r)}$ und $\sigma_{12}^{(r)}$. Bei einer einfache Stichprobe vom Umfang $n = 10$ wurden die folgenden Werte beobachtet:[156]

i	Y_{i1}	Y_{i2}
1	0,52	2,04
2	-0,35	0,64
3	-1,80	1,51
4	0,95	4,91
5	-0,51	2,88
6	1,23	1,04
7	0,94	?
8	0,29	?
9	-1,43	?
10	1,36	?

[156] Die Merkmalsträger wurden anhand der Vollständigkeit der Daten geordnet.

Der Parameter der Bernoulli-verteilten Zufallsvariable R_2 ist durch die Gleichung (6.59) zu schätzen:

$$\hat{\varepsilon}_1 = \frac{n-q}{n} = 0{,}4 \tag{6.86}$$

Die Parameterschätzer, die sich ausschließlich auf beobachtete Daten beziehen, können unmittelbar aus (6.60) - (6.63) bestimmt werden:

$$\hat{\mu}_1^{(0)} = \bar{y}_1^{(0)} = 0{,}01 \qquad \hat{\mu}_2^{(0)} = \bar{y}_2^{(0)} = 2{,}17 \qquad \hat{\mu}_1^{(1)} = \bar{y}_1^{(1)} = 0{,}29$$

$$\hat{\sigma}_{11}^{(0)} = 1{,}05 \qquad \hat{\sigma}_{22}^{(0)} = 2{,}01 \qquad \hat{\sigma}_{12}^{(0)} = 0{,}42$$

$$\hat{\sigma}_{11}^{(1)} = 1{,}13 \qquad \hat{\sigma}_{11} = 1{,}10 \tag{6.87}$$

Unter der MAR-Annahme und der daraus abgeleiteten Restriktion (6.68) erhält man die folgenden drei Schätzwerte für $\mu_2^{(1)}, \sigma_{22}^{(1)}$ und $\sigma_{12}^{(1)}$:[157]

$$\begin{aligned}\hat{\mu}_2^{(1)} &= \bar{y}_2^{(0)} + \hat{b}^{(0)}\left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right) \\ &= \bar{y}_2^{(0)} + \frac{\hat{\sigma}_{12}^{(0)}}{\hat{\sigma}_{11}^{(0)}}\left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right) \\ &= 2{,}28\end{aligned} \tag{6.88}$$

$$\begin{aligned}\hat{\sigma}_{22}^{(1)} &= \hat{\sigma}_{22}^{(0)} + \left(\frac{\hat{\sigma}_{12}^{(0)}}{\hat{\sigma}_{11}^{(0)}}\right)^2 \left(\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11}^{(0)}\right) \\ &= 2{,}02\end{aligned} \tag{6.89}$$

$$\hat{\sigma}_{12}^{(1)} = \frac{\hat{\sigma}_{12}^{(0)}}{\hat{\sigma}_{11}^{(0)}} \hat{\sigma}_{11}^{(1)} = 0{,}45 \tag{6.90}$$

Der Schätzwert für den Erwartungswert von Y_2 ist, wie in (6.74) nachgewiesen wurde, durch folgende Gleichung bestimmbar:

$$\hat{\mu}_2 = \bar{y}_2^{(0)} + \hat{\varepsilon}_1 \frac{\hat{\sigma}_{12}^{(0)}}{\hat{\sigma}_{12}^{(0)}}\left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right) = 2{,}21 \tag{6.91}$$

[157] Vgl. Formeln (6.70)-(6.72).

Die Schätzwerte für die Varianz von Y_2 und die Kovarianz von Y_1 und Y_2 sind:[158]

$$\hat{\sigma}_{22} = \hat{\sigma}_{22}^{(0)} + \left(\frac{\hat{\sigma}_{12}^{(0)}}{\hat{\sigma}_{11}^{(0)}}\right)^2 \left(\hat{\sigma}_{11} - \hat{\sigma}_{11}^{(0)}\right) = 2{,}02 \tag{6.92}$$

$$\hat{\sigma}_{12} = \frac{\hat{\sigma}_{12}^{(0)}}{\hat{\sigma}_{11}^{(0)}} \hat{\sigma}_{11} = 0{,}44 \tag{6.93}$$

Wird von einem MNAR-Ausfallmechanismus und der Restriktion (6.79) ausgegangen, so erhält man die folgenden Schätzwerte für die Parameter von $P_\omega\left(Y_2^{(1)}\right)$:[159]

$$\hat{\mu}_2^{(1)} = \bar{y}_2^{(0)} + \frac{\left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right)}{\hat{d}^{(0)}} = \bar{y}_2^{(0)} + \frac{\left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right)\hat{\sigma}_{22}^{(0)}}{\hat{\sigma}_{12}^{(0)}} = 3{,}50 \tag{6.94}$$

$$\hat{\sigma}_{22}^{(1)} = \hat{\sigma}_{22}^{(0)} + \frac{\left(\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11}^{(0)}\right)\left(\hat{\sigma}_{22}^{(0)}\right)^2}{\left(\hat{\sigma}_{12}^{(0)}\right)^2} = 3{,}82 \tag{6.95}$$

$$\hat{\sigma}_{12}^{(1)} = \hat{\sigma}_{12}^{(0)} + \frac{\left(\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11}^{(0)}\right)\hat{\sigma}_{22}^{(0)}}{\hat{\sigma}_{12}^{(0)}} = 0{,}80 \tag{6.96}$$

Die Schätzer des ersten und zweiten Moments von Y_2 sowie der Kovarianz von Y_1 und Y_2 sind:[160]

$$\hat{\mu}_2 = \bar{y}_2^{(0)} + \hat{\varepsilon}_1 \frac{\hat{\sigma}_{22}^{(0)}}{\hat{\sigma}_{12}^{(0)}} \left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right) = 2{,}70 \tag{6.97}$$

$$\hat{\sigma}_{22} = \hat{\sigma}_{22}^{(0)} + \frac{\left(\hat{\sigma}_{11} - \hat{\sigma}_{11}^{(0)}\right)\left(\hat{\sigma}_{22}^{(0)}\right)^2}{\left(\hat{\sigma}_{12}^{(0)}\right)^2} = 3{,}14 \tag{6.98}$$

$$\hat{\sigma}_{12} = \hat{\sigma}_{12}^{(0)} + \frac{\left(\hat{\sigma}_{11} - \hat{\sigma}_{11}^{(0)}\right)\hat{\sigma}_{11}^{(0)}}{\hat{\sigma}_{12}^{(0)}} = 0{,}66 \tag{6.99}$$

Die Parameterschätzwerte bei Vorliegen von MNAR weichen erheblich von den Werten in (6.88)-(6.93), die unter dem MAR-Ausfallmechanismus ermittelt wurden, ab. Diese Unterschiede verdeutlichen, welche Auswirkungen die verschiedenen Me-

---

[158] Vgl. Formeln (6.75) und (6.76).
[159] Die Gleichungen wurden in (6.80)-(6.82) hergeleitet.
[160] Vgl. Herleitung in (6.83)-(6.85).

chanismen besitzen können, so dass eine Sensitivitätsanalyse bezüglich der Schätzungen in Betracht zu ziehen ist. Die folgenden Ausführungen zeigen, wie diese Analyse im bivariaten Pattern-Mixture Modell realisiert werden kann.

6.4.2.4 Sensitivitätsanalyse im bivariaten Pattern-Mixture Modell

In dem vorherigen Kapitel wurde gezeigt, dass bei der Erfüllung der bedingten Unabhängigkeitsannahme

$$R_2 \perp\!\!\!\perp Y_1 \mid Y_2 \tag{6.100}$$

die Parameter des Pattern-Mixture Modells auch bei einem Datenausfall vom Typ MNAR unverzerrt geschätzt werden können. Ein Ansatz, der die Unabhängigkeitsannahme in (6.100) nicht voraussetzt und damit weniger restriktiv ist, fügt in dem Modell zur Erklärung des Ausfallmechanismus eine Variable Y_2^* hinzu. Die Zufallsvariable Y_2^* setzt sich dabei aus der Summe von Y_1 und dem λ-fachen von Y_2 zusammen:[161]

$$Y_2^* = Y_1 + \lambda Y_2 \tag{6.101}$$

Es sei angenommen, dass der Datenausfall von Y_2 nur von Y_2^* direkt abhängig ist. Um die Zusammenhänge zwischen den betrachteten Variablen in anschaulicher Weise zu verdeutlichen, können diese im Rahmen der Graphentheorie modelliert werden.[162] Die folgende Abbildung zeigt die beiden gerichteten Graphen (I)* und (II)*, die sich aus den Modellannahmen ergeben.

[161] Vgl. Little (1994), S. 476.
[162] Vgl. Lauritzen (1996); Voß (2000), S. 693ff.

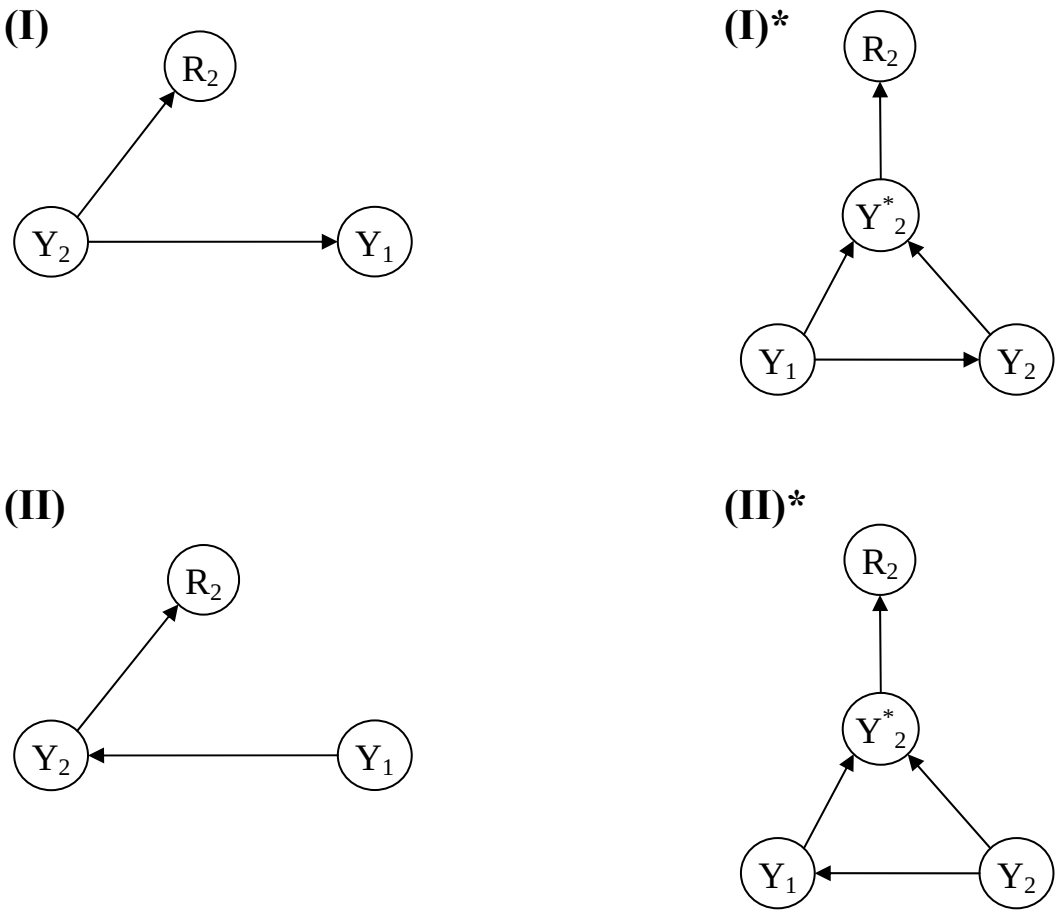


Abbildung 6.4: Graphische Darstellung des MNAR-Mechanismus aus Beispiel 6.8 (linke Abbildungen) und unter Berücksichtigung der Variable Y_2^* (rechte Abbildungen)

Der Parameter λ fungiert in dem Ansatz als Einflussfaktor von Y_2 auf den Datenausfall: Im Spezialfall $\lambda = 0$ hängt das Fehlen von Y_2 nur von Y_1 ab und der Ausfallmechanismus ist vom Typ MAR. Falls $\lambda = 1$ ist, wird der MNAR-Ausfallmechanismus durch die Summe von Y_1 und Y_2 vollständig erklärt. Für $\lambda \to \infty$ ist Y_2 die einzige erklärende Variable für die (Nicht-)Beobachtung von Y_2 und es gilt die bedingte Unabhängigkeitsbeziehung aus (6.100), so dass der in Beispiel 6.8 angenommene MNAR-Mechanismus ebenfalls durch den Ansatz dargestellt werden kann.[163] Durch die Berücksichtigung von Y_2^* können somit verschiedene Konstellationen des Daten-

[163] Vgl. Little (1994), S. 476.

ausfalls vom Typ MNAR modelliert werden. Darüber hinaus erlaubt diese Modellierung eine Sensitivitätsbetrachtung, bei der unter Annahme verschiedener Werte für λ der Einfluss des Ausfallmechanismus auf die Parameterschätzungen untersucht wird.

Aufgrund des graphentheoretischen Kriteriums der d-Separation[164] kann aus beiden Graphen (I)* und (II)* in Abbildung 6.4 die bedingte Unabhängigkeit von R und Y_1 gegeben Y_2^* $\left(Y_1 \perp\!\!\!\perp R_2 \mid Y_2^*\right)$ abgeleitet werden. Demzufolge ist die bedingte Verteilung von Y_1 gegeben Y_2^* in beiden Klassen gleich:[165]

$$P_{\omega^*_{0,1\cdot 2}}(Y_1 \mid Y_2^*, R_2 = 0) = P_{\omega^*_{1,1\cdot 2}}(Y_1 \mid Y_2^*, R_2 = 1) \tag{6.102}$$

Die Parametervektoren $\omega^*_{0,1\cdot 2} = \left(c^{(0)*}, d^{(0)*}, \sigma^{(0)*}_{11\cdot 2}\right)$ und $\omega^*_{1,1\cdot 2} = \left(c^{(1)*}, d^{(1)*}, \sigma^{(1)*}_{11\cdot 2}\right)$, für die aufgrund von (6.102)

$$\omega^*_{0,1\cdot 2} = \omega^*_{1,1\cdot 2} \tag{6.103}$$

gilt, beziehen sich auf die folgenden Regressionsmodelle:

$$\begin{aligned} Y_1^{(r)} &= c^{(r)*} + d^{(r)*} Y_2^{(r)*} + W^{(r)} \\ &= c^{(r)*} + d^{(r)*}\left(Y_1^{(r)} + \lambda Y_2^{(r)}\right) + W^{(r)} \qquad (r = 0{,}1) \end{aligned}$$

Dabei ist die Zufallsvariable $Y_2^{(r)*}$ durch

$$Y_2^{(r)*} := Y_2^* \mid R_2 = r \;\; (r = 0,\, 1)$$

definiert, während die Variable $W^{(r)}$ den Erwartungswert $E\left(W^{(r)}\right) = 0$ und die Varianz $\mathrm{Var}\left(W^{(r)}\right) = \sigma^{(r)*}_{11\cdot 2}$ besitzt.

Aus der Annahme (6.103) folgt $c^{(0)*} = c^{(1)*}$ und $d^{(0)*} = d^{(1)*}$, so dass für den Erwartungswert von $Y_2^{(1)}$ gilt:[166]

$$\mu_2^{(1)} = \frac{\mu_1^{(1)}\left(1 - d^{(1)*}\right) - c^{(1)*}}{\lambda\, d^{(1)*}} = \frac{\mu_1^{(1)}\left(1 - d^{(0)*}\right) - c^{(0)*}}{\lambda\, d^{(0)*}} \tag{6.104}$$

---

[164] Vgl. Lauritzen (1996), S. 48f.
[165] Vgl. Little (1994), S. 476.
[166] Vgl. Little (1994), S. 476.

Der Regressionsparameter $d^{(0)*}$ kann in folgender Weise durch die Varianzen von $Y_1^{(0)}$ und $Y_2^{(0)}$ und der Kovarianz dieser beiden Zufallsvariablen hergeleitet werden:

$$\begin{aligned} d^{(0)*} &= \frac{\operatorname{cov}\left(Y_1^{(0)}, Y_2^{(0)*}\right)}{\operatorname{Var}\left(Y_2^{(0)*}\right)} \\ &= \frac{\operatorname{cov}\left(Y_1^{(0)}, Y_1^{(0)} + \lambda Y_2^{(0)}\right)}{\operatorname{Var}\left(Y_1^{(0)} + \lambda Y_2^{(0)}\right)} \\ &= \frac{\operatorname{Var}\left(Y_1^{(0)}\right) + \lambda \operatorname{cov}\left(Y_1^{(0)}, Y_2^{(0)}\right)}{\operatorname{Var}\left(Y_1^{(0)}\right) + \lambda^2 \operatorname{Var}\left(Y_2^{(0)}\right) + 2\lambda \operatorname{cov}\left(Y_1^{(0)}, Y_2^{(0)}\right)} \\ &= \frac{\sigma_{11}^{(0)} + \lambda \sigma_{12}^{(0)}}{\sigma_{11}^{(0)} + \lambda^2 \sigma_{22}^{(0)} + 2\lambda \sigma_{12}^{(0)}} \end{aligned} \tag{6.105}$$

Durch Bildung der Schätzer für $\mu_1^{(1)}$, $d^{(0)*}$ und $c^{(0)*}$ in (6.104) erhält man

$$\begin{aligned} \hat{\mu}_2^{(1)} &= \frac{\bar{y}_1^{(1)}\left(1 - \hat{d}^{(0)*}\right) - \hat{c}^{(0)*}}{\lambda \hat{d}^{(0)*}} \\ &= \frac{\bar{y}_1^{(1)}\left(1 - \hat{d}^{(0)*}\right) - \left(\bar{y}_1^{(0)} - \hat{d}^{(0)*}\left(\bar{y}_1^{(0)} + \lambda \bar{y}_2^{(0)}\right)\right)}{\lambda \hat{d}^{(0)*}} \\ &= \bar{y}_2^{(0)} + \left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right)\frac{1 - \hat{d}^{(0)*}}{\lambda \hat{d}^{(0)*}}. \end{aligned} \tag{6.106}$$

Für den Schätzer von $\mu_2^{(1)}$ gilt aufgrund der Umformung in (6.105):

$$\hat{\mu}_2^{(1)} = \bar{y}_2^{(0)} + \hat{\nu}\left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right) \qquad \left(\hat{\nu} = \frac{\lambda \hat{\sigma}_{22}^{(0)} + \hat{\sigma}_{12}^{(0)}}{\lambda \hat{\sigma}_{12}^{(0)} + \hat{\sigma}_{11}^{(0)}}\right) \tag{6.107}$$

Der Erwartungswert von Y_2 wird durch

$$\begin{aligned} \hat{\mu}_2 &= \left(1 - \hat{\varepsilon}_1\right)\hat{\mu}_2^{(0)} + \hat{\varepsilon}_1 \hat{\mu}_2^{(1)} \\ &= \frac{q}{n}\bar{y}_2^{(0)} + \frac{n-q}{n}\left(\bar{y}_2^{(0)} + \hat{\nu}\left(\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right)\right) \\ &= \bar{y}_2^{(0)} + \hat{\nu}\left(\left(\bar{y}_1 - \frac{q}{n}\bar{y}_1^{(0)}\right) - \frac{n-q}{n}\bar{y}_1^{(0)}\right) \\ &= \bar{y}_2^{(0)} + \hat{\nu}\left(\bar{y}_1 - \bar{y}_1^{(0)}\right) \end{aligned} \tag{6.108}$$

geschätzt.

Weiterhin kann gezeigt werden, dass

$$\hat{\sigma}_{22}^{(1)} = \hat{\sigma}_{22}^{(0)} + (\hat{\nu})^2 \left(\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11}^{(0)}\right) \tag{6.109}$$

und

$$\hat{\sigma}_{12}^{(1)} = \hat{\sigma}_{12}^{(0)} + \hat{\nu}\left(\hat{\sigma}_{11}^{(1)} - \hat{\sigma}_{11}^{(0)}\right) \tag{6.110}$$

mit dem in (6.107) definierten Schätzer $\hat{\nu}$ ist und somit analog für die weiteren Parameterschätzer bezüglich der gesamten Daten gilt:[167]

$$\hat{\sigma}_{22} = \hat{\sigma}_{22}^{(0)} + (\hat{\nu})^2 \left(\hat{\sigma}_{11} - \hat{\sigma}_{11}^{(0)}\right) \tag{6.111}$$

$$\hat{\sigma}_{12} = \hat{\sigma}_{12}^{(0)} + \hat{\nu}\left(\hat{\sigma}_{11} - \hat{\sigma}_{11}^{(0)}\right) \tag{6.112}$$

Beispiel 6.9:

Gegeben seien die beobachteten Werte für Y_1 und Y_2 aus Beispiel 6.8. In Abhängigkeit von λ ergeben sich die folgenden Schätzwerte für die Parameter in dem unvollständig beobachteten Pattern $r = 1$:

$$\hat{\mu}_2^{(1)} = 2{,}17 + \frac{\lambda 2{,}01 + 0{,}42}{\lambda 0{,}42 + 1{,}05}(0{,}29 - 0{,}01) = \frac{1{,}47\lambda + 2{,}40}{0{,}42\lambda + 1{,}05}$$

$$\hat{\sigma}_{22}^{(1)} = 2{,}01 + \left(\frac{\lambda 2{,}01 + 0{,}42}{\lambda 0{,}42 + 1{,}05}\right)^2 (1{,}13 - 1{,}05) = 2{,}01 + \left(\frac{0{,}57\lambda + 0{,}12}{0{,}42\lambda + 1{,}05}\right)^2$$

$$\hat{\sigma}_{12}^{(1)} = 0{,}42 + \frac{\lambda 2{,}01 + 0{,}42}{\lambda 0{,}42 + 1{,}05}(1{,}13 - 1{,}05) = \frac{0{,}34\lambda + 0{,}47}{0{,}42\lambda + 1{,}05}$$

[167] Vgl. Little/Rubin (2002), S. 334.

Für den Erwartungswert und die Varianz von Y_2 sowie die Kovarianz von Y_1 und Y_2 in den gesamten Daten resultieren die folgenden Schätzwerte:

$$\hat{\mu}_2 = 2{,}17 + \frac{\lambda 2{,}01 + 0{,}42}{\lambda 0{,}42 + 1{,}05}(0{,}12 - 0{,}01) = \frac{1{,}13\lambda + 2{,}33}{0{,}42\lambda + 1{,}05}$$

$$\hat{\sigma}_{22} = 2{,}01 + \left(\frac{\lambda 2{,}01 + 0{,}42}{\lambda 0{,}42 + 1{,}05}\right)^2 (1{,}10 - 1{,}05) = 2{,}01 + \left(\frac{0{,}45\lambda + 0{,}09}{0{,}42\lambda + 1{,}05}\right)^2$$

$$\hat{\sigma}_{12} = 0{,}42 + \left(\frac{\lambda 2{,}01 + 0{,}42}{\lambda 0{,}42 + 1{,}05}\right)(1{,}10 - 1{,}05) = \frac{0{,}28\lambda + 0{,}46}{0{,}42\lambda + 1{,}05}$$

Unter der Annahme, dass $\lambda \geq 0$ ist und somit Y_1 und Y_2 positiv korrelieren, kann der Wertebereich für den Schätzer $\hat{\mu}_2$ eingeschränkt werden. Die – in Abhängigkeit von λ ermittelten – verschiedenen Schätzwerte für μ_2 sind in dem folgenden Diagramm dargestellt.

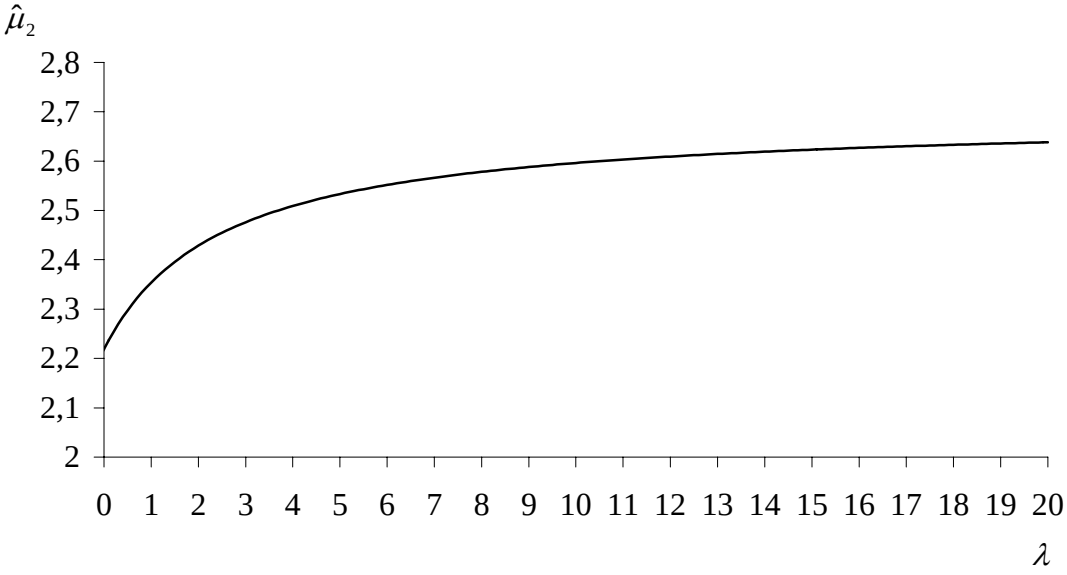


Abbildung 6.5: Schätzwerte für μ_2 in Abhängigkeit von λ (Beispiel 6.9)

Der Schätzer für den Erwartungswert von Y_2 liegt unter der Restriktion $\lambda \geq 0$ im Intervall [2,21; 2,70]. Die Untergrenze des Intervalls resultiert dabei aus $\lambda = 0$, wodurch dieser Wert dem in Beispiel 6.8 ermittelten Schätzwert entspricht, welcher bei Vorliegen von MAR bestimmt wurde.[168] Wird hingegen der MNAR-

[168] Vgl. Schätzwert für μ_2 in (6.91).

Ausfallmechanismus unter der Restriktion $\omega_{0,1\cdot 2} = \omega_{1,1\cdot 2}$ angenommen, so korrespondiert der Schätzwert mit der Obergrenze des Intervalls.[169] Der Parameter λ, welcher in diesem Beispiel Werte zwischen Null und unendlich annehmen kann, ist als ein Index für die Nicht-Ignorierbarkeit des Ausfallmechanismus zu interpretieren:[170] Größere Werte für λ implizieren einen starken direkten Einfluss von Y_2 auf den Datenausfall, während Werte für λ nahe Null auf eine hinreichende Erklärung des Ausfalls durch die Kovariate Y_1 hindeuten. Kann durch zusätzliches Wissen der Wertebereich für λ weiter begrenzt werden, so führt dies entsprechend zu einer geringeren Schwankungsbreite des Schätzers $\hat{\mu}_2$.

Weitere Eigenschaften des Pattern-Mixture Modells ergeben sich, indem Y_1 und Y_2 mit Hilfe der ersten und zweiten Momente von $Y_1^{(0)}$ bzw. $Y_2^{(0)}$ standardisiert werden. Ausgehend von der Gleichung für den Schätzer des Erwartungswerts in (6.108) ist

$$\begin{aligned}
\frac{\hat{\mu}_2 - \bar{y}_2^{(0)}}{\sqrt{\hat{\sigma}_{22}^{(0)}}} &= \frac{\hat{\nu}}{\sqrt{\hat{\sigma}_{22}^{(0)}}}\left(\bar{y}_1 - \bar{y}_1^{(0)}\right) \\
&= \frac{\lambda\hat{\sigma}_{22}^{(0)} + \hat{\sigma}_{12}^{(0)}}{\sqrt{\hat{\sigma}_{22}^{(0)}}\left(\lambda\hat{\sigma}_{12}^{(0)} + \hat{\sigma}_{12}^{(0)}\right)}\left(\frac{q}{n}\bar{y}_1^{(0)} + \frac{n-q}{n}\bar{y}_1^{(1)} - \bar{y}_1^{(0)}\right) \\
&= \frac{\sqrt{\hat{\sigma}_{11}^{(0)}}\left(\lambda\hat{\sigma}_{22}^{(0)} + \hat{\sigma}_{12}^{(0)}\right)}{\sqrt{\hat{\sigma}_{22}^{(0)}}\left(\lambda\hat{\sigma}_{12}^{(0)} + \hat{\sigma}_{12}^{(0)}\right)}\left(\frac{n-q}{n}\right)\left(\frac{\bar{y}_1^{(1)} - \bar{y}_1^{(0)}}{\sqrt{\hat{\sigma}_{11}^{(0)}}}\right) \\
&= \frac{\left(\lambda\frac{\sqrt{\hat{\sigma}_{22}^{(0)}}}{\sqrt{\hat{\sigma}_{11}^{(0)}}} + \frac{\hat{\sigma}_{12}^{(0)}}{\sqrt{\hat{\sigma}_{11}^{(0)}\hat{\sigma}_{22}^{(0)}}}\right)}{\left(\lambda\frac{\hat{\sigma}_{12}^{(0)}}{\hat{\sigma}_{11}^{(0)}} + 1\right)}\left(\frac{n-q}{n}\right)\left(\frac{\bar{y}_1^{(1)} - \bar{y}_1^{(0)}}{\sqrt{\hat{\sigma}_{11}^{(0)}}}\right) \\
&= \left(\frac{\lambda^* + \hat{\rho}_{12}^{(0)}}{\lambda^*\hat{\rho}_{12}^{(0)} + 1}\right)\left(\frac{n-q}{n}\right)\left(\frac{\bar{y}_1^{(1)} - \bar{y}_1^{(0)}}{\sqrt{\hat{\sigma}_{11}^{(0)}}}\right)
\end{aligned} \tag{6.113}$$

mit dem geschätzten Korrelationskoeffizienten innerhalb der vollständig beobachteten Daten

[169] Vgl. Schätzwert für μ_2 in (6.97).
[170] Vgl. Little/Rubin (2002), S. 334.

$$\hat{\rho}_{12}^{(0)} = \frac{\hat{\sigma}_{12}^{(0)}}{\sqrt{\hat{\sigma}_{11}^{(0)}\hat{\sigma}_{22}^{(0)}}}$$

und

$$\lambda^* = \sqrt{\frac{\hat{\sigma}_{22}^{(0)}}{\hat{\sigma}_{11}^{(0)}}}\lambda \,.^{171}$$

Aus Gleichung (6.113) ist ersichtlich, dass der standardisierte Mittelwert von Y_2 in den gesamten Datensätzen

$$\frac{\hat{\mu}_2 - \bar{y}_2^{(0)}}{\sqrt{\hat{\sigma}_{22}^{(0)}}}$$

proportional zum standardisierten Mittelwert von Y_1 in den unvollständigen Datensätzen

$$\frac{\bar{y}_1^{(1)} - \bar{y}_1^{(0)}}{\sqrt{\hat{\sigma}_{11}^{(0)}}}$$

ist. Weiterhin führt ein hoher Anteil fehlender Werte

$$\frac{n-q}{n}$$

zu einer hohen Gewichtung des standardisierten Mittelwertes von $Y_1^{(1)}$ bezüglich der Schätzung in (6.113): Die vollständig vorliegenden Informationen über die Zufallsvariable Y_1 werden in hohem Maße zur Schätzung des Erwartungswertes von Y_2 verwendet. Der Schätzer $\hat{\mu}_2$ wird außerdem durch den Koeffizienten

$$\frac{\lambda^* + \hat{\rho}_{12}^{(0)}}{\lambda^* \hat{\rho}_{12}^{(0)} + 1}$$

beeinflusst. Falls vorausgesetzt werden kann, dass der Korrelationskoeffizient $\hat{\rho}_{12}^{(0)}$ positiv sowie $\lambda^* \geq 0$ ist, so gilt für diesen Koeffizienten:[172]

$$\hat{\rho}_{12}^{(0)} \leq \frac{\lambda^* + \hat{\rho}_{12}^{(0)}}{\lambda^* \hat{\rho}_{12}^{(0)} + 1} \leq \frac{1}{\hat{\rho}_{12}^{(0)}} \tag{6.114}$$

[171] Vgl. Little (1994), S. 477.
[172] Vgl. Little (1994), S. 477.

Die Unter- und Obergrenze resultieren aus den beiden Extremfällen $\lambda^* = 0$ bzw. $\lambda^* = \infty$. Die Breite des Intervalls hängt allein von der Korrelation in den vollständig beobachteten Daten ab: Ist $\hat{\rho}_{12}^{(0)}$ annähernd Null, so ist der Wertebereich für den Koeffizienten in (6.114) sehr groß und die weiteren Annahmen für λ^* sind von entscheidender Bedeutung für die Schätzung von μ_2. Falls hingegen ein starker positiver, linearer Zusammenhang zwischen $Y_1^{(0)}$ und $Y_2^{(0)}$ besteht, ist der Wertebereich des Koeffizienten deutlich eingeschränkt und die weiteren Festlegungen bezüglich λ^* sind weniger relevant für $\hat{\mu}_2$. Dies verdeutlicht, dass der Bias von Schätzungen in hohem Maße beschränkt werden kann, wenn die Kovariate stark mit der Variable in den vollständig beobachteten Daten korreliert, die vom Ausfall betroffen ist.[173]

Im Fall einer geringen Korrelation sind die Annahmen bezüglich λ^* problematisch: Aus den beobachteten Daten können keine Rückschlüsse für λ^* gezogen werden, so dass die Sensitivität der Parameterschätzungen anhand verschiedener Werte von λ^* zu untersuchen ist. Little (1994) weist darauf hin, dass diese Problematik nicht als Nachteil des Pattern-Mixture Modells zu sehen ist, sondern allgemein auf das Fehlen von Y_2 zurückgeführt werden kann. Durch Selection Modelle wird dieses Problem nur scheinbar umgangen, da die Parameterschätzungen in hohem Maße von den Verteilungsannahmen sowohl für $\underline{Y}$ als auch für $\underline{R}$ gegeben $\underline{y}$ abhängig sind, die anhand der verfügbaren Daten nicht überprüft werden können.[174]

Die bisherigen Ausführungen in Kapitel 6.4.2 bezogen sich auf zwei Zufallsvariablen, von denen lediglich eine Variable fehlende Werte in einer Stichprobe aufweist. Wird letztere Beschränkung aufgehoben, so können in einer Stichprobe insgesamt vier verschiedene Pattern auftreten, wie die folgende Darstellung zeigt.

[173] Vgl. Little (1994), S. 477.
[174] Vgl. Little (1994), S. 477; Stolzenberg/Relles (1990), S. 406ff.

i	R_{i1}	R_{i2}	Pattern r
1	0	0	0
2	0	0	0
3	0	1	1
4	1	0	2
		.	
		.	
		.	
n	1	1	3

Abbildung 6.6: Anzahl der verschiedenen Pattern im Pattern-Mixture Modell bei $k = 2$ Zufallsvariablen

Die Parametervektoren $\omega_1,\ldots, \omega_3$ der bedingten Verteilungen $P_{\omega_r}\left(Y_1, Y_2 \mid R = r\right)$ ($r = 1,\ldots,3$) können nur aufgrund von Annahmen (Identifying Restrictions) geschätzt werden. Diese Restriktionen können für jedes Pattern unterschiedlich festgelegt werden, wodurch es möglich ist, den Fall des kompletten Datenausfalls (Pattern $r = 3$) separat von der partiellen Nichtbeobachtung (Pattern $r = 1, 2$) zu behandeln. Im nächsten Kapitel wird insbesondere gezeigt, wie unter diesem Aspekt eine Sensitivitätsbetrachtung der Parameterschätzung erfolgen kann.

6.4.3 Pattern-Set Mixture Modelle

Eine differenziert zu betrachtende Problematik im Kontext der Behandlung von fehlenden Werten ist die vollständige Antwortverweigerung (Unit Nonresponse). In vielen Fällen erscheint es einerseits plausibel, dass das vollständige Fehlen von Datensätzen auf die nicht beobachteten Werte zurückzuführen und somit der Ausfall durch einen Mechanismus vom Typ MNAR zu modellieren ist. Auf der anderen Seite kann für das Fehlen einzelner Werte (Item Nonresponse) häufig ein anderer Mechanismus (z.B. MAR) unterstellt werden. Pattern-Set Mixture Modelle beziehen sich auf diese Problemstellung, indem bei diesen Ansätzen Unit Nonresponse und Item Nonresponse gesondert behandelt werden. Diese innerhalb der Verfahren vorgenommene strikte Trennung wird durch die zwei verschieden definierten Response-

Indikatorvariablen deutlich: Der Ausfall von Datensatz *i* sei durch eine Zufallsvariable

$$R_i^{(1)} = \begin{cases} 0 & \text{Unit Response} \\ 1 & \text{Unit Nonresponse} \end{cases} \qquad (i = 1,\ldots, n)$$

mit den Realisationen $r_i^{(1)}$ beschrieben, während $\underline{r}_i^{(2)} = \left(r_{i1}^{(2)},...,r_{ik}^{(2)}\right)$ die Realisationen einer *k*-dimensionalen Zufallsvariablen $\underline{R}_i^{(2)} = (R_{i1}^{(2)},...,R_{ik}^{(2)})$ sind, die das Fehlen einzelner Werte im Datensatz *i* mit

$$R_{ij}^{(2)} = \begin{cases} 0 & \text{Beobachtung von } y_{ij} \text{ (Item Response)} \\ 1 & \text{Datenausfall von } y_{ij} \text{ (Item Nonresponse)} \end{cases} \qquad (i = 1,\ldots, n; j = 1,\ldots,k)$$

indizieren. In Analogie zur Notation in Kapitel 2 werden die *n* Beobachtungen von $R_i^{(1)}$ und $\underline{R}_i^{(2)}$ in den Vektoren $\mathbf{R}^{(1)}$ bzw. $\mathbf{R}^{(2)}$ zusammengefasst.

Die gemeinsame Verteilung von den Response-Indikatoren und den beobachteten und fehlenden Variablen in Datensatz *i* wird im Pattern-Set Mixture Modell durch

$$P_{\omega,\psi,\varepsilon}(\underline{Y}_i, R_i^{(1)}, \underline{R}_i^{(2)}) = P_\omega(\underline{Y}_i \mid R_i^{(1)}) P_\psi(\underline{R}_i^{(2)} \mid \underline{Y}_i, R_i^{(1)}) P_\varepsilon(R_i^{(1)}) \tag{6.115}$$

faktorisiert.[175] Im stetigen Fall bilden die bedingten Dichten von $\underline{Y}_i$ gegeben dem Unit Nonresponse-Indikator $R_i^{(1)} = r^{(1)}$ ($r^{(1)} = 0, 1$) sowie die Verteilung $P_\varepsilon(R_i^{(1)})$ ein Pattern-Mixture Modell, das unter der Annahme einer multivariaten Normalverteilung von $\underline{Y}$ in den beiden Pattern mit $\omega = \left(\omega_0, \omega_1\right) = \left((\mu^{(0)}, \Sigma^{(0)}), (\mu^{(1)}, \Sigma^{(1)})\right)$ wie folgt definiert ist:

$$\begin{aligned} &(1)\quad \left(\underline{Y}_i \mid R_i^{(1)} = r^{(1)}\right) \sim N\left(\mu^{(r^{(1)})}, \Sigma^{(r^{(1)})}\right) && r^{(1)} = 0{,}1 \;;\; i = (1,...,n) \\ &&& \mu^{(r^{(1)})} = \left(\mu_1^{(r^{(1)})},...,\mu_k^{(r^{(1)})}\right) \\ &(2)\quad R_i^{(1)} \sim B\left(1, \varepsilon_1\right) && \varepsilon = \left(\varepsilon_0, \varepsilon_1\right), \varepsilon_0 = 1 - \varepsilon_1 \end{aligned} \tag{6.116}$$

[175] Vgl. Little/Rubin (2002), S.314.

Für die ersten *q* Datensätze sei mindestens ein Wert beobachtet worden (Unit Response), während bei den restlichen *n-q* Merkmalsträgern jegliche Werte fehlen (Unit Nonresponse).[176]

Für die gemeinsame Verteilung der gesamten Daten und der beiden Indikatorvariablen gilt unter Beachtung der iid-Eigenschaft der Stichprobe:

$$P_{\omega,\psi,\varepsilon}(\mathbf{Y},\mathbf{R}^{(1)},\mathbf{R}^{(2)}) =$$

$$\prod_{i=1}^{n} P_{\omega_{r_i^{(1)}}}(\underline{Y}_i = \underline{y}_i \mid R_i^{(1)} = r_i^{(1)}) P_{\psi}(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_i = \underline{y}_i, R_i^{(1)} = r_i^{(1)}) P_{\varepsilon_{r_i^{(1)}}}(R_i^{(1)} = r_i^{(1)})$$

$$= \prod_{i=1}^{q} P_{\omega_0}(\underline{Y}_i = \underline{y}_i \mid R_i^{(1)} = 0)\, P_{\psi}(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_i = \underline{y}_i, R_i^{(1)} = 0) P_{\varepsilon_0}(R_i^{(1)} = 0)$$

$$\cdot \prod_{i=q+1}^{n} P_{\omega_1}(\underline{Y}_i = \underline{y}_i \mid R_i^{(1)} = 1) P_{\psi}(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_i = \underline{y}_i, R_i^{(1)} = 1) P_{\varepsilon_1}(R_i^{(1)} = 1)$$

$$= (1-\varepsilon_1)^q \prod_{i=1}^{q} P_{\omega_0}(\underline{Y}_i = \underline{y}_i \mid R_i^{(1)} = 0)\, P_{\psi}(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_i = \underline{y}_i, R_i^{(1)} = 0)$$

$$\cdot (\varepsilon_1)^{n-q} \prod_{i=q+1}^{n} P_{\omega_1}(\underline{Y}_i = \underline{y}_i \mid R_i^{(1)} = 1) P_{\psi}(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_i = \underline{y}_i, R_i^{(1)} = 1) \qquad (6.117)$$

Innerhalb der partiell und vollständig beobachteten Datensätze (Unit Response) sind $P_{\omega_0}(\underline{Y}_i = \underline{y}_i \mid R_i^{(1)} = 0)$ und $P_{\psi}(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_i = \underline{y}_i, R_i^{(1)} = 0)$ die beiden Faktoren eines Selection Modells:

$$P_{\omega_0,\psi}(\underline{Y}_i = \underline{y}_i, \underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid R_i^{(1)} = 0) =$$
$$P_{\omega_0}(\underline{Y}_i = \underline{y}_i \mid R_i^{(1)} = 0) P_{\psi}(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_i = \underline{y}_i, R_i^{(1)} = 0) \qquad (6.118)$$

Weiterhin folgt aus $R_i^{(1)} = 1$ (Unit Nonresponse) unmittelbar $R_{ij}^{(2)} = 1$ für $j = 1,...,k$ bzw. $\underline{R}_i^{(2)} = (1,1,...,1)$, so dass die bedingte Wahrscheinlichkeit

$$P_{\psi}(\underline{R}_i^{(2)} = 1 \mid \underline{Y}_i = \underline{y}_i, R_i^{(1)} = 1) = 1 \qquad (i = 1,\ldots,n) \qquad (6.119)$$

und

$$\underline{y}_{mis,i} = \underline{y}_i \qquad \forall i = q+1,....,n \qquad (6.120)$$

ist.

---

[176] O.B.d.A. können die Datensätze bezüglich des Auftretens von Unit Nonresponse geordnet werden.

Hängt der Ausfall einzelner Werte nicht von den Ausprägungen selbst ab, dann gilt die MAR-Annahme für den Fall von Item Nonresponse. Damit ist die bedingte Wahrscheinlichkeit des Responseindikators $\underline{R}_i^{(2)} = \underline{r}_i^{(2)}$ in den unvollständig beobachteten Datensätzen unabhängig von den fehlenden Werten $\underline{y}_{mis,i}$ gegeben $\underline{y}_{obs,i}$:[177]

$$P_\psi(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_i = \underline{y}_i, R_i^{(1)} = 0) = P_\psi(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_{obs,i} = \underline{y}_{obs,i}, R_i^{(1)} = 0) \qquad (6.121)$$

Unter Einbeziehung von (6.119)-(6.121) erhält man die Likelihood-Funktion der beobachteten Daten – gegeben ω, ψ und ε - durch Integration der gemeinsamen Verteilung in (6.117) über die fehlenden Werte $\mathbf{y}_{mis}$:

$$\begin{aligned}
&L(\mathbf{y}_{obs}, \mathbf{r}^{(1)}, \mathbf{r}^{(2)} \mid \omega, \psi, \varepsilon) = \\
&\int (1-\varepsilon_1)^q \prod_{i=1}^{q} P_{\omega_0}(\underline{Y}_i = \underline{y}_i \mid R_i^{(1)} = 0) \cdot P_\psi(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_i = \underline{y}_i, R_i^{(1)} = 0) \\
&\cdot (\varepsilon_1)^{n-q} \prod_{i=q+1}^{n} P_{\omega_1}(\underline{Y}_i = \underline{y}_i \mid R_i^{(1)} = 1)\, d\mathbf{y}_{mis} \\
&= \int (1-\varepsilon_1)^q \prod_{i=1}^{q} P_{\omega_0}(\underline{Y}_i = \underline{y}_i \mid R_i^{(1)} = 0) \cdot P_\psi(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_{obs,i} = \underline{y}_{obs,i}, R_i^{(1)} = 0) \\
&\cdot (\varepsilon_1)^{n-q} \prod_{i=q+1}^{n} P_{\omega_1}(\underline{Y}_{mis,i} = \underline{y}_{mis,i} \mid R_i^{(1)} = 1)\, d\mathbf{y}_{mis} \\
&= (1-\varepsilon_1)^q \prod_{i=1}^{q} P_{\omega_0}(\underline{Y}_{obs,i} = \underline{y}_{obs,i} \mid R_i^{(1)} = 0) \cdot P_\psi(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_{obs,i} = \underline{y}_{obs,i}, R_i^{(1)} = 0) \\
&\cdot (\varepsilon_1)^{n-q} \prod_{i=q+1}^{n} \int P_{\omega_1}(\underline{Y}_{mis,i} = \underline{y}_{mis,i} \mid R_i^{(1)} = 1)\, d\,\underline{y}_{mis,i} \\
&= (1-\varepsilon_1)^q \prod_{i=1}^{q} P_{\omega_0}(\underline{Y}_{obs,i} = \underline{y}_{obs,i} \mid R_i^{(1)} = 0)\, P_\psi(\underline{R}_i^{(2)} = \underline{r}_i^{(2)} \mid \underline{Y}_{obs,i} = \underline{y}_{obs,i}, R_i^{(1)} = 0)(\varepsilon_1)^{n-q}
\end{aligned} \qquad (6.122)$$

Die Parametervektoren ω_0 und ψ des Selection Modells in (6.118) sowie der Parameter ε_1 des in (6.116) definierten Pattern-Mixture Modells können durch Maximierung der Likelihood-Funktion geschätzt werden. Ist der Mechanismus für Unit Nonresponse vom Typ MNAR und somit nicht ignorierbar, so sind Annahmen bezüglich des nicht bestimmbaren Parametervektors ω_1 der bedingten Dichte

$$P_{\omega_1}(\underline{Y}_i = \underline{y}_i \mid R_i^{(1)} = 1)$$

[177] Vgl. Little/Rubin (2002), S. 315.

zu treffen. Dies kann zum einen durch die im letzten Kapitel eingeführten Identifying Restrictions erfolgen. Eine weitere, einfachere Vorgehensweise besteht in der Festlegung, dass der Schätzwert des Parametervektors ω_1 bei MNAR durch eine (prozentuale) Änderung λ gegenüber dem Wert bei Ignorierbarkeit des Ausfallmechanismus ausgedrückt wird. Durch Variierung von λ kann wiederum die Sensitivität des Parameterschätzers der marginalen Verteilung von $\underline{Y}$ überprüft werden.[178] Das folgende Beispiel zeigt, dass diese Vorgehensweise auch auf kategoriale Daten übertragen werden kann.

Beispiel 6.10:

Es werden zwei binäre Zufallsvariablen Y_1 und Y_2 betrachtet, deren Realisationen in einer Stichprobe vom Umfang n unvollständig beobachtet wurden. Anhand der Pattern können die beobachteten Daten in vier Kontingenztabellen zusammengefasst werden, wobei $n_{ab}^{(r)}$ die absolute Häufigkeit von $Y_1 = a$ und $Y_2 = b$ in Pattern r ist ($a = 0,1$; $a = 0,1$):

Pattern 0 (vollständige Beobachtung von Y_1 und Y_2):

	Y_2=0	Y_2=1	
Y_1=0	$n_{00}^{(0)} = 40$	$n_{01}^{(0)} = 20$	$n_{0\bullet}^{(0)} = 60$
Y_1=1	$n_{10}^{(0)} = 25$	$n_{11}^{(0)} = 15$	$n_{1\bullet}^{(0)} = 40$
	$n_{\bullet 0}^{(0)} = 65$	$n_{\bullet 1}^{(0)} = 35$	$n^{(0)} = 100$

Pattern 1 (ausschließliche Beobachtung von Y_1):

	Y_2=0	Y_2=1	
Y_1=0	$n_{00}^{(1)} = ?$	$n_{01}^{(1)} = ?$	$n_{0\bullet}^{(1)} = 20$
Y_1=1	$n_{10}^{(1)} = ?$	$n_{11}^{(1)} = ?$	$n_{1\bullet}^{(1)} = 80$
	$n_{\bullet 0}^{(1)} = ?$	$n_{\bullet 1}^{(1)} = ?$	$n^{(1)} = 100$

Pattern 2 (ausschließliche Beobachtung von Y_2):

	Y_2=0	Y_2=1	
Y_1=0	$n_{00}^{(2)} = ?$	$n_{01}^{(2)} = ?$	$n_{0\bullet}^{(2)} = ?$
Y_1=1	$n_{10}^{(2)} = ?$	$n_{11}^{(2)} = ?$	$n_{1\bullet}^{(2)} = ?$
	$n_{\bullet 0}^{(2)} = 50$	$n_{\bullet 1}^{(2)} = 150$	$n^{(2)} = 200$

Pattern 3 (Y_1 und Y_2 nicht beobachtet):

	Y_2=0	Y_2=1	
Y_1=0	$n_{00}^{(3)} = ?$	$n_{01}^{(3)} = ?$	$n_{0\bullet}^{(3)} = ?$
Y_1=1	$n_{10}^{(3)} = ?$	$n_{11}^{(3)} = ?$	$n_{1\bullet}^{(3)} = ?$
	$n_{\bullet 0}^{(3)} = ?$	$n_{\bullet 1}^{(3)} = ?$	$n^{(3)} = 50$

178 Vgl. Allison (2002), S. 83f. Allison setzte die beschriebene Vorgehensweise anhand eines Beispiels mit mehrfacher Ersetzung der fehlenden Werte um.

Die Daten in dem Pattern r (r = 0,...,3) werden durch die absoluten Häufigkeiten $n_{00}^{(r)}$, $n_{01}^{(r)}, n_{10}^{(r)}$ und $n_{11}^{(r)}$ vollständig beschrieben. Diese Häufigkeiten seien Realisationen der mehrdimensionalen Zufallsvariablen $\left(N_{00}^{(r)}, N_{01}^{(r)}, N_{10}^{(r)}, N_{11}^{(r)}\right)$ (r = 0,...,3), die multinomialverteilt sind mit den Parametern $n^{(r)}$ und $\omega^{(r)} = \left(\omega_{00}^{(r)}, \omega_{01}^{(r)}, \omega_{10}^{(r)}, \omega_{11}^{(r)}\right)$:

$$\left(N_{00}^{(r)}, N_{01}^{(r)}, N_{10}^{(r)}, N_{11}^{(r)}\right) \sim M\left(n^{(r)}, \omega_{00}^{(r)}, \omega_{01}^{(r)}, \omega_{10}^{(r)}, \omega_{11}^{(r)}\right) \qquad (r = 0,...,3)$$

Aufgrund der vollständigen Beobachtung von Y_1 und Y_2 im Pattern r = 0 können die Schätzwerte für den Parametervektor ω_0 unmittelbar durch

$$\hat{\omega}_{ab}^{(0)} = \frac{n_{ab}^{(0)}}{n^{(0)}} \qquad (a = 0,1;\ b = 0,1)$$

bestimmt werden:

Pattern 0:

	Y_2=0	Y_2=1	
Y_1=0	$\hat{\omega}_{00}^{(0)} = 0{,}4$	$\hat{\omega}_{01}^{(0)} = 0{,}2$	$\hat{\omega}_{0\bullet}^{(0)} = 0{,}6$
Y_1=1	$\hat{\omega}_{10}^{(0)} = 0{,}25$	$\hat{\omega}_{11}^{(0)} = 0{,}15$	$\hat{\omega}_{1\bullet}^{(0)} = 0{,}4$
	$\hat{\omega}_{\bullet 0}^{(0)} = 0{,}65$	$\hat{\omega}_{\bullet 1}^{(0)} = 0{,}35$	

Für den Fall von Item Nonresponse wird ein ignorierbarer Ausfallmechanismus vom Typ MAR angenommen, so dass unter anderem

$$P(Y_{i2} = b \mid Y_{i1} = a, R_i^{(1)} = 0, \underline{R}_i^{(2)} = 1) = P(Y_{i2} = b \mid Y_{i1} = a, R_i^{(1)} = 0, \underline{R}_i^{(2)} = 0)$$

bzw.

$$\frac{P(Y_{i1} = a, Y_{i2} = b, R_i^{(1)} = 0, \underline{R}_i^{(2)} = 0)}{P(Y_{i1} = a, R_i^{(1)} = 0, \underline{R}_i^{(2)} = 0)} = \frac{P(Y_1 = a, Y_2 = b, R_i^{(1)} = 0, \underline{R}_i^{(2)} = 1)}{P(Y_1 = a, R_i^{(1)} = 0, \underline{R}_i^{(2)} = 1)} \qquad (6.123)$$

für i = 1,...,n gilt.

Die Parameter $\omega_{ab}^{(r)}$ (a = 0,1; b = 0,1; r = 0,...,2) der Multinomialverteilungen ent-

sprechen den bedingten Wahrscheinlichkeiten $P(Y_{i1} = a, Y_{i2} = b \mid R_i^{(1)} = 0, \underline{R}_i^{(2)} = r)$ für $i = 1,...,n$. Beispielsweise ist bei Nichtbeobachtung von Y_{i2} die bedingte Wahrscheinlichkeit von $Y_{i1} = 0$ und $Y_{i2} = 0$ im Pattern $r = 1$ dem Parameter $\omega_{00}^{(1)}$ gleichzusetzen:

$$P(Y_{i1} = 0, Y_{i2} = 0 \mid R_i^{(1)} = 0, \underline{R}_i^{(2)} = 1) = \omega_{00}^{(1)} \qquad (i = 1,...,n)$$

Unter Einbeziehung von Gleichung (6.123) lässt sich diese bedingte Wahrscheinlichkeit durch

$$\begin{aligned} P(Y_{i1} = 0, Y_{i2} = 0 \mid R_i^{(1)} = 0, \underline{R}_i^{(2)} = 1) &= \frac{P(Y_{i1} = 0, Y_{i2} = 0, R_i^{(1)} = 0, \underline{R}_i^{(2)} = 1)}{P(R_i^{(1)} = 0, \underline{R}_i^{(2)} = 1)} \\ &= \frac{P(Y_{i1} = 0, Y_{i2} = 0, R_i^{(1)} = 0, \underline{R}_i^{(2)} = 0)}{P(Y_{i1} = 0, R_i^{(1)} = 0, \underline{R}_i^{(2)} = 0)} \frac{P(Y_{i1} = 0, R_i^{(1)} = 0, \underline{R}_i^{(2)} = 1)}{P(R_i^{(1)} = 0, \underline{R}_i^{(2)} = 1)} \end{aligned} \qquad (6.124)$$

ermitteln, und der Schätzer für den Parameter $\omega_{00}^{(1)}$ der Multinomialverteilung in Pattern $r = 1$ ist aus den bekannten absoluten (Rand-)Häufigkeiten bestimmbar:

$$\hat{\omega}_{00}^{(1)} = \frac{n_{00}^{(0)}}{n_{0\bullet}^{(0)}} \frac{n_{0\bullet}^{(1)}}{n^{(1)}} = 0{,}13$$

Analog können die restlichen Parameter in dem Vektor ω_1 geschätzt werden:

Pattern 1:

	Y_2=0	Y_2=1	
Y_1=0	$\hat{\omega}_{00}^{(1)} = 0{,}13$	$\hat{\omega}_{01}^{(1)} = 0{,}07$	$\hat{\omega}_{0\bullet}^{(1)} = 0{,}2$
Y_1=1	$\hat{\omega}_{10}^{(1)} = 0{,}5$	$\hat{\omega}_{11}^{(1)} = 0{,}3$	$\hat{\omega}_{1\bullet}^{(1)} = 0{,}8$
	$\hat{\omega}_{\bullet 0}^{(1)} = 0{,}63$	$\hat{\omega}_{\bullet 1}^{(1)} = 0{,}37$	

Im Pattern $r = 2$ wurde lediglich die Zufallsvariable Y_1 nicht beobachtet (Item Nonresponse), und für den Schätzer des Parameters

$$\omega_{00}^{(2)} = P(Y_{i1} = 0, Y_{i2} = 0 \mid R_i^{(1)} = 0, \underline{R}_i^{(2)} = 2)$$

gilt unter der MAR-Annahme:[179]

$$\hat{\omega}_{00}^{(2)} = \frac{n_{00}^{(0)}}{n_{\bullet 0}^{(0)}} \frac{n_{\bullet 0}^{(2)}}{n^{(2)}} = 0{,}15$$

Die weiteren Schätzer der Parameter in dem Pattern $r = 2$ sind in der folgenden Kontingenztabelle zusammengefasst:

Pattern 2:

	Y_2=0	Y_2=1	
Y_1=0	$\hat{\omega}_{00}^{(2)} = 0{,}15$	$\hat{\omega}_{01}^{(2)} = 0{,}43$	$\hat{\omega}_{0\bullet}^{(2)} = 0{,}58$
Y_1=1	$\hat{\omega}_{10}^{(2)} = 0{,}1$	$\hat{\omega}_{11}^{(2)} = 0{,}32$	$\hat{\omega}_{1\bullet}^{(2)} = 0{,}42$
	$\hat{\omega}_{\bullet 0}^{(2)} = 0{,}25$	$\hat{\omega}_{\bullet 1}^{(2)} = 0{,}75$	

Im Pattern $r = 3$ wurde keine der beiden Zufallsvariablen beobachtet (Unit Nonresponse). Da im Pattern $r = 1$ $\hat{\omega}_{1\bullet}^{(1)} > \hat{\omega}_{0\bullet}^{(1)}$ und im Pattern $r = 2$ $\hat{\omega}_{\bullet 1}^{(2)} > \hat{\omega}_{\bullet 0}^{(2)}$ gilt, erscheint die Annahme plausibel, dass hohe Werte von Y_1 *und* Y_2 ein häufiger Grund für Unit Nonresponse sind. Dementsprechend wird im Folgenden angenommen, dass der Parameter $\omega_{11}^{(3)}$ im Pattern $r = 3$ um λ höher ist als der entsprechende Parameter $\omega_{11}^{(1,2)}$ in den beiden *zusammengefassten* Pattern 1 und 2:

$$\omega_{11}^{(3)} = \omega_{11}^{(1,2)} + \lambda \qquad (\lambda > 0) \tag{6.125}$$

Der Schätzer für $\omega_{11}^{(1,2)}$ ist dabei ein gewichtetes arithmetisches Mittel aus $\hat{\omega}_{11}^{(1)}$ und $\hat{\omega}_{11}^{(2)}$:

$$\hat{\omega}_{11}^{(1,2)} = \frac{n_{11}^{(1)} + n_{11}^{(2)}}{n^{(1)} + n^{(2)}} = \frac{n^{(1)}\hat{\omega}_{11}^{(1)} + n^{(2)}\hat{\omega}_{11}^{(2)}}{n^{(1)} + n^{(2)}} = 0{,}31$$

Um die Parameter des Modells schätzen zu können, sind neben (6.125) weitere Annahmen zu treffen. Aus diesem Grund wird im Folgenden angenommen, dass die

[179] Die Herleitung erfolgt analog zu den Gleichungen (6.123) und (6.124).

Wahrscheinlichkeiten für unterschiedliche Werte von Y_1 und Y_2 ($Y_1 = 0,\ Y_2 = 1$ bzw. $Y_1 = 1,\ Y_2 = 0$) bei Unit Nonresponse und Item Nonresponse identisch sind:

$$\omega_{01}^{(3)} = \omega_{01}^{(1,2)} \quad , \quad \omega_{10}^{(3)} = \omega_{10}^{(1,2)} \tag{6.126}$$

Wegen

$$\sum_{a=0}^{1}\sum_{b=0}^{1}\omega_{ab}^{(1,2)} = 1 \ , \ \sum_{a=0}^{1}\sum_{b=0}^{1}\omega_{ab}^{(3)} = 1$$

gilt unter diesen Annahmen für den Parameter $\omega_{00}^{(3)}$:

$$\omega_{00}^{(3)} = 1 - \omega_{11}^{(3)} - \omega_{01}^{(3)} - \omega_{10}^{(3)} = 1 - \left(\omega_{11}^{(1,2)} + \lambda\right) - \omega_{01}^{(1,2)} - \omega_{10}^{(1,2)} = \omega_{00}^{(1,2)} - \lambda$$

Somit ergeben sich im Pattern $r = 3$ die folgenden von λ $(0 < \lambda \leq 0{,}15)$ abhängigen Parameterschätzwerte $\hat{\omega}_{ab}^{(3)}$ ($a = 0,1,\ b = 0,1$):[180]

Pattern 3:

	Y_2=0	Y_2=1	
Y_1=0	$\hat{\omega}_{00}^{(3)} = 0{,}15 - \lambda$	$\hat{\omega}_{01}^{(3)} = 0{,}31$	$\hat{\omega}_{0\bullet}^{(3)} = 0{,}46 - \lambda$
Y_1=1	$\hat{\omega}_{10}^{(3)} = 0{,}23$	$\hat{\omega}_{11}^{(3)} = 0{,}31 + \lambda$	$\hat{\omega}_{1\bullet}^{(3)} = 0{,}54 + \lambda$
	$\hat{\omega}_{\bullet 0}^{(3)} = 0{,}38 - \lambda$	$\hat{\omega}_{\bullet 1}^{(3)} = 0{,}62 + \lambda$	

Aus den Schätzern in den einzelnen Pattern r ($r = 0,...,3$) können ebenfalls die Parameter der gemeinsamen Verteilung $P(Y_1, Y_2)$ geschätzt werden. Die Herleitung erfolgt über die mit n_{ab} bezeichneten absoluten Häufigkeiten von $Y_1 = a$ und $Y_2 = b$ ($a = 0,1,\ b = 0,1$), welche Realisationen der multinomialverteilten Zufallsvariablen $\left(N_{00}, N_{01}, N_{10}, N_{11}\right)$ mit den Parametern n und $\theta = \left(\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11}\right)$ sind:

$$\left(N_{00}, N_{01}, N_{10}, N_{11}\right) \sim M\left(n, \theta_{00}, \theta_{01}, \theta_{10}, \theta_{11}\right)$$

[180] Die Restriktion $0 < \lambda \leq 0{,}15$ folgt unmittelbar aus der Bedingung $0 \leq \hat{\omega}_{00}^{(3)} \leq 1$

Die Parameter

$$\theta_{ab} = P_{\theta_{ab}}(Y_1 = a, Y_2 = b) \quad (a = 0,1; b = 0,1)$$

der Multinomialverteilung können aus dem gewichteten arithmetischen Mittel von $\hat{\omega}_{ab}^{(0)},...,\hat{\omega}_{ab}^{(3)}$ geschätzt werden:

$$\hat{\theta}_{ab} = \frac{1}{n}\sum_{r=0}^{3} n_{ab}^{(r)} = \frac{1}{n}\sum_{r=0}^{3} n^{(r)}\hat{\omega}_{ab}^{(r)} \qquad (a = 0,1; b = 0,1)$$

Für den Parameter θ_{11} erhält man somit als Schätzwert:

$$\hat{\theta}_{11} = \frac{1}{n}\sum_{r=0}^{3} n_{11}^{(r)} = \frac{1}{n}\sum_{r=0}^{3} n^{(r)}\hat{\omega}_{11}^{(r)} = 0,28 + 0,11\lambda$$

Die folgende Kontingenztabelle enthält die weiteren Parameterschätzwerte, die sich im Pattern-Set Mixture Modell für die gesamten Daten ergeben:

Pattern 0-3 (gesamte Daten):

	Y_2=0	Y_2=1	
Y_1=0	$\hat{\theta}_{00} = 0,20 - 0,11\lambda$	$\hat{\theta}_{01} = 0,28$	$\hat{\theta}_{0\bullet} = 0,48 - 0,11\lambda$
Y_1=1	$\hat{\theta}_{10} = 0,24$	$\hat{\theta}_{11} = 0,28 + 0,11\lambda$	$\hat{\theta}_{1\bullet} = 0,52 + 0,11\lambda$
	$\hat{\theta}_{\bullet 0} = 0,44 - 0,11\lambda$	$\hat{\theta}_{\bullet 1} = 0,56 + 0,11\lambda$	

Aufgrund der aus dem Pattern $r = 3$ resultierenden Restriktion $0 < \lambda \leq 0,15$ und des geringen Anteils von Unit Nonresponse am gesamten Stichprobenumfang sind die Auswirkungen von λ auf die Schätzungen begrenzt. Sind die Annahmen in dem Modell erfüllt, so erhält man die folgenden Bereiche für die Schätzer von θ_{00} und θ_{11}:

$$0,18 \leq \hat{\theta}_{00} < 0,2 \quad , \quad 0,28 < \hat{\theta}_{11} \leq 0,3$$

6.4.4 Fazit

Pattern-Mixture Modelle stellen im Rahmen der Behandlung von fehlenden Werten eine vielversprechende Alternative zu den in Kapitel 6.3 diskutierten Selection Modellen dar. Dies wird auch durch die neuere statistische Literatur zu diesem Themen-

gebiet belegt, die sich eingehend mit den Pattern-Mixture Modellen befasst.[181] Ein wesentlicher methodischer Vorteil der Pattern-Mixture Modelle ist die strikte Trennung der Parameter anhand ihrer Identifizierbarkeit aus den Daten. Für die nicht identifizierbaren Parameter können Restriktionen (Identifying Restrictions) eingeführt werden, die sich im Allgemeinen auf den Zusammenhang zu den identifizierbaren Parametern beziehen. Im Gegensatz zu den Selection Modellen ist es somit in unkomplizierter Weise möglich, die Sensitivität der Schätzungen durch Variierung dieser Restriktionen zu beurteilen.[182]

Im Rahmen von Pattern-Set Mixture Modellen können die beiden verschiedenen Ansätze (Selection und Pattern-Mixture) kombiniert werden, um verschiedene Ausfallmechanismen bezüglich des vollständigen Datenausfalls (Unit Nonresponse) und des Fehlens einzelner Werte (Item Nonresponse) zu berücksichtigen. In Kapitel 6.4.3 wurde für den diskreten Fall dargestellt, wie eine Parameterschätzung erfolgen kann, falls die MAR-Annahme bei Item Nonresponse gilt und ein nicht ignorierbarer Ausfallmechanismus bei Unit Nonresponse vorliegt.

Werden mehrere Zufallsvariablen in einem Pattern-Mixture Modell betrachtet, so besteht im Allgemeinen die Problematik, dass eine Vielzahl von Parametern nicht identifizierbar ist.[183] Um in diesem Fall eine Parameterschätzung oder Sensitivitätsbetrachtung zu ermöglichen, sind die Annahmen bezüglich des nicht ignorierbaren Ausfallmechanismus häufig sehr restriktiv zu formulieren. Dies stellt jedoch keinen Nachteil gegenüber den Selection Modellen dar, da in diesen Modellen die Problematik durch die zugrunde liegenden Verteilungsannahmen nur scheinbar umgangen wird. In dem folgenden Kapitel wird u.a. dieser Aspekt durch einen methodischen Vergleich der beiden Ansätze verdeutlicht.

6.5 Vergleich von Selection und Pattern-Mixture Modellen

Die Behandlung von fehlenden Werten durch die in der statistischen Literatur erwähnten Methoden erfordert generell die Festlegung von Annahmen, wenn der Aus-

---

[181] Vgl. Ekholm/Skinner (1998); Molenberghs et al. (1998); Forster/Smith (1998); Daniels/Hogan (2000); Kenward et al. (2003); Storck et al. (2000).
[182] Vgl. Kapitel 6.4.2.4 sowie Beispiel 6.10 in Kapitel 6.4.3.
[183] Dies trifft insbesondere bei Unit Nonresponse zu, da in dem entsprechenden Pattern alle Parameter nicht identifizierbar sind (vgl. Beispiel 6.10).

fallmechanismus nicht ignorierbar ist. Pattern-Mixture und Selection Modelle basieren dabei, wie in den vorangegangenen Kapiteln dargestellt, auf unterschiedlichen Verteilungsannahmen. Während in Selection Modellen die Verteilungstypen von $P_\theta(\underline{Y})$ und $P_\psi(\underline{R} \mid \underline{Y})$ zu spezifizieren sind, ist im Pattern-Mixture Modell der Verteilungstyp von $P_\omega(\underline{Y} \mid R = r)$ für jedes Pattern r ($r = 0,\ldots, s$) vorzugeben.[184]

Um die beiden Ansätze methodisch miteinander zu vergleichen, können die bedingte Verteilung $P_\omega(\underline{Y} \mid \underline{R})$ und die Verteilung $P_\varepsilon(\underline{R})$ des Pattern-Mixture Modells

$$P_{\omega,\varepsilon}(\underline{Y}, \underline{R}) = P_\omega(\underline{Y} \mid \underline{R}) P_\varepsilon(\underline{R})$$

aus dem Selection Modell

$$P_{\theta,\psi}(\underline{Y}, \underline{R}) = P_\theta(\underline{Y})\, P_\psi(\underline{R} \mid \underline{Y})$$

mit den Parametern θ und ψ bestimmt werden:[185, 186]

$$P_{\theta,\psi}(\underline{R}) = \int_{-\infty}^{\infty} P_\psi(\underline{R} \mid \underline{Y}) P_\theta(\underline{Y}) d\,\underline{y} \tag{6.127}$$

$$P_{\theta,\psi}(\underline{Y} \mid \underline{R}) = \frac{P_\psi(\underline{R} \mid \underline{Y}) P_\theta(\underline{Y})}{P_{\theta,\psi}(\underline{R})} \tag{6.128}$$

In letzterem Ansatz ist die bedingte Verteilung $P(\underline{Y} \mid \underline{R})$ sowohl von θ als auch von ψ abhängig, da ein bestimmter Verteilungstyp von $\underline{Y}$ mit dem Parameter θ anzunehmen ist und der Ausfallmechanismus mit dem Parameter ψ nur durch Einbeziehung von $\underline{Y}$ modelliert werden kann. Hingegen wird im Pattern-Mixture Modell vorausgesetzt, dass die bedingte Verteilung $P(\underline{Y} \mid \underline{R})$ – und *nicht* die marginale Verteilung von $\underline{Y}$ – durch das jeweilige Element des Parametervektors $\omega = (\omega_0,\ldots,\omega_s)$ angegeben werden kann.

Bei dem Selection Modell ist die Bestimmung der marginalen Verteilung $P(\underline{R})$ in (6.127) nur durch Integration (bzw. Summation) über die teilweise unbeobachteten Werte von $\underline{Y}$ möglich, und somit ist – neben dem Parameter ψ – die Kenntnis des

---

[184] In vielen Fällen wird im Pattern-Mixture Modell angenommen, dass der gleiche Verteilungstyp in allen Pattern vorliegt (vgl. bivariates Pattern Mixture Modell in Kapitel 6.4.2).
[185] Vgl. Holland, P. W. (1986), S. 149.
[186] Im diskreten Fall ist das Integrationszeichen durch ein Summenzeichen zu ersetzen.

Parameters θ der Verteilung von $\underline{Y}$ erforderlich. Dagegen gilt a priori im Pattern-Mixture Modell, dass die diskrete Zufallsvariable $\underline{R}$ die nur vom Parametervektor ε abhängige Verteilung $P_{\varepsilon}(\underline{R})$ besitzt und der Zusammenhang zwischen $\underline{R}$ und $\underline{Y}$ einzig durch $P_{\omega}(\underline{Y} \mid \underline{R})$ erklärt wird.[187]

In Pattern-Mixture Modellen geht durch die Aufteilung der Daten nach vollständiger Beobachtung und unvollständiger Beobachtung eindeutig hervor, dass ohne weitere Annahmen fehlende Werte nicht plausibel ersetzt werden können. Im univariaten Fall ist die bedingte Verteilung $P_{\omega_1}(\underline{Y} \mid \underline{R} = 1)$ vollkommen unbekannt, und es sind sowohl Annahmen bezüglich des Verteilungstyps als auch in Bezug auf den Parametervektor ω_1 zu treffen, um z.B. die Momente von $\underline{Y}$ schätzen zu können.[188] Bei Selection Modellen sind diese Restriktionen weniger offensichtlich, jedoch zeigt sich durch Zerlegung der marginalen Verteilung von $\underline{Y}$

$$P_{\theta}(\underline{Y}) = \left(1 - P_{\theta,\psi}(\underline{R} = 1)\right) P_{\theta,\psi}(\underline{Y} \mid \underline{R} = 0) + P_{\theta,\psi}(\underline{R} = 1) P_{\theta,\psi}(\underline{Y} \mid \underline{R} = 1)$$

die Bedeutung der unbekannten bedingten Verteilung $P_{\theta,\psi}(\underline{Y} \mid \underline{R} = 1)$ und der ebenfalls nicht ermittelbaren Ausfallwahrscheinlichkeit $P_{\theta,\psi}(\underline{R} = 1)$ für die Schätzungen in diesem Ansatz.[189] Diese Unkenntnis wird durch die Annahmen bezüglich der Verteilungstypen von $P_{\theta}(\underline{Y})$ und $P_{\psi}(\underline{R} \mid \underline{Y})$ umgangen. Simulationsstudien von Glynn et al. (1986) belegen allerdings, dass die Fehlspezifikation der Verteilungstypen im Selection Modell zu erheblich verzerrten Parameterschätzungen führt, während sich die Pattern-Mixture Modellierung demgegenüber als relativ robust erweist.[190] Insofern ist auch eine Präferenz der Selection Modelle gegenüber den Pattern-Mixture Modellen, wie sie derzeit bei der Behandlung von fehlenden Werten unter MNAR immer noch zu konstatieren ist, nicht berechtigt.

Die grundsätzlich verschiedenen Annahmen der Methoden führen im Allgemeinen zu unterschiedlichen Ersetzungen von fehlenden Werten, sofern die beiden Modelle

---

[187] Vgl. Kapitel 6.4.1.
[188] Vgl. Beispiel 6.7. In diesem Beispiel wurde eine binäre Zufallsvariable betrachtet, so dass der Verteilungstyp (Binomialverteilung) bereits vorgegeben war.
[189] Vgl. Holland (1986), S. 150.
[190] Vgl. Glynn et al. (1986), S. 127ff., Die Relevanz der korrekten Spezifikation von Verteilungstypen im Selection Modell wurde in weiteren Simulationen bestätigt (vgl. Stolzenberg/Relles (1990)).

zur Parameterschätzung verwendet werden. Das folgende Beispiel soll diesen Aspekt verdeutlichen.

Beispiel 6.11:

Es wird das folgende Pattern-Mixture Modell für eine stetige eindimensionale Zufallsvariable $\underline{Y}$ betrachtet:

$$\underline{Y} \mid R_1 = r \sim N\left(\mu^{(r)}, \sigma^{(r)}\right) \qquad r = 0,1$$

$$R_1 \sim B(1, \varepsilon_1)$$

Im Weiteren sollen die Parameter $\omega = \left(\mu^{(0)}, \sigma^{(0)}, \mu^{(1)}, \sigma^{(1)}\right)$ und $\varepsilon = \left(\varepsilon_0, \varepsilon_1\right)$ $\left(\varepsilon_0 = 1 - \varepsilon_1\right)$ des Modells aus simulierten Daten geschätzt werden. Die Simulation erfolgte dabei durch die folgenden beiden Schritte:

1. Ziehen eines Wertes q der binomialverteilten Zufallsvariable

   $$Q \sim B\left(n, (1 - \varepsilon_1)\right) \quad ; n = 1.000, \varepsilon_1 = 0,25$$

2. Ziehen von q Werten $\underline{y}_1,\ldots,\underline{y}_q$ der standardnormalverteilten Zufallsvariablen

   $$\underline{Y}_1,...,\underline{Y}_q \sim N(0,1)$$

Durch diese Vorgehensweise werden unter den Parametervorgaben

$$\mu^{(0)} = 0\,,\ \sigma^{(0)} = 1\,,\ \varepsilon_1 = 0,25$$

die beobachteten Daten (Pattern $r = 0$) bei einem Stichprobenumfang von $n = 1.000$ simuliert. Wird für den (nicht identifizierbaren) Parameter $\sigma^{(1)}$ die Restriktion

$$\sigma^{(0)} = \sigma^{(1)}$$

angenommen, so ist die Dichtefunktion von $\underline{Y}$ in Abhängigkeit vom unbekannten Parameter $\mu^{(1)}$ darstellbar. Die folgende Abbildung stellt beispielhaft für $\mu^{(1)} = 4$ und den aus den Daten geschätzten Parametern

$$\hat{\mu}^{(0)} = 0,02\ ,\hat{\sigma}^{(0)} = 0,98\,,\ \hat{\varepsilon}_1 = 0,24$$

die bedingten Dichten $P_{\hat{\omega}}(\underline{Y} \mid \underline{R} = r)$ $(r = 0,1)$ sowie die Dichte $P_{\hat{\omega},\hat{\varepsilon}}(\underline{Y})$ graphisch dar.

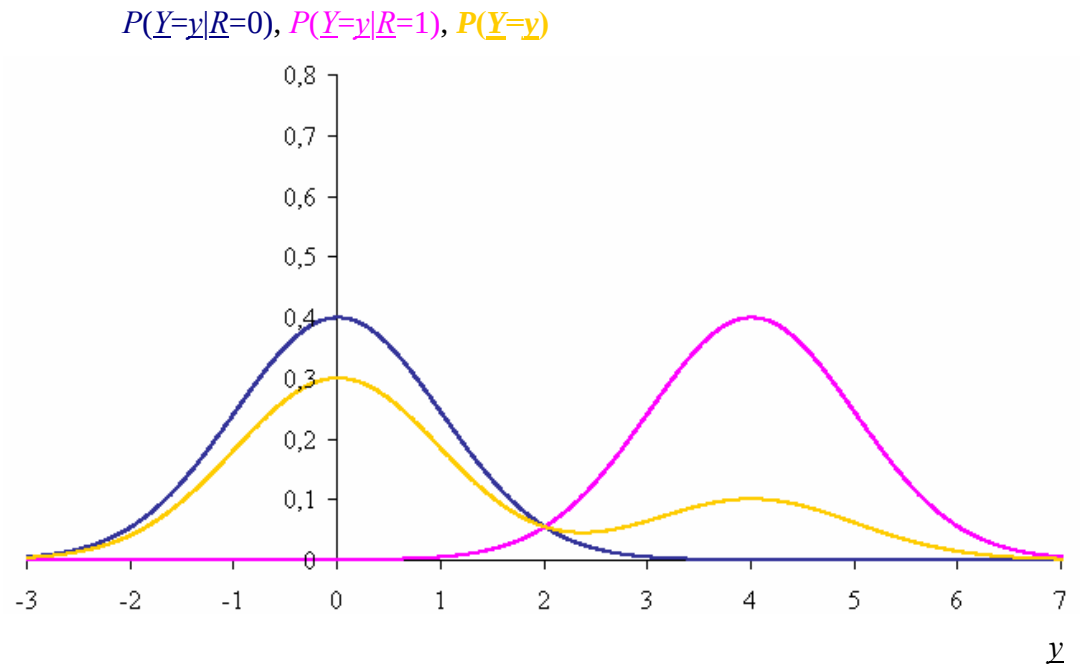


Abbildung 6.7: bedingte Dichten $P_{\hat{\omega}}(\underline{Y} \mid \underline{R} = r)$ $(r = 0,1)$ und daraus resultierende Dichtefunktion von $\underline{Y}$ im Pattern-Mixture Modell

Die Dichtefunktion von $\underline{Y}$ ist eine Mischung von zwei Normalverteilungen, die durch die Parametervektoren ω und ε beschrieben wird. Demzufolge ist die Selection Modellierung in diesem Beispiel nicht gerechtfertigt, da diese eine durch θ beschriebene Verteilung von $\underline{Y}$ voraussetzt. Im Folgenden soll dargestellt werden, welche Auswirkungen auf die Parameterschätzungen sich aus den Annahmen des Selection Modells, die in diesem Beispiel nicht erfüllt sind, ergeben. Im Selection Ansatz sei $\underline{Y}$ annahmegemäß normalverteilt mit den Parametern μ und σ:

$$\underline{Y} \sim N\left(\mu, \sigma^2\right)$$

Für den Ausfallmechanismus wird das in Kapitel 6.3.2.2 beschriebene Schwellenwert-Modell

$$P_{\psi}\left(\underline{R} = 1 \mid \underline{Y} \geq 0\right) = \psi$$

$$P_{\psi}\left(\underline{R} = 1 \mid \underline{Y} < 0\right) = 0$$

angenommen. Aus den beobachteten Daten $\underline{y}_1,\ldots,\underline{y}_q$ der Stichprobe werden die Parameter μ, σ und ψ durch

$$\hat{\mu} = 0{,}2\ , \hat{\sigma} = 1{,}1\ , \hat{\psi} = 0{,}43$$

geschätzt, und man erhält die folgenden Dichtefunktionen aus der Selection Modellierung:[191]

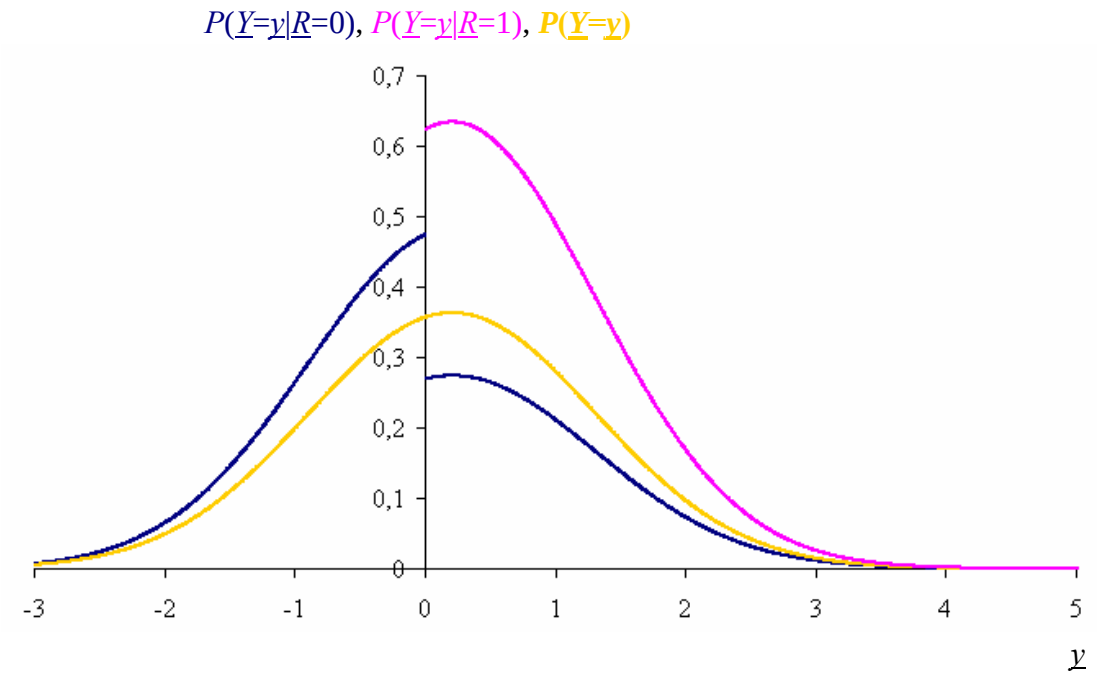


Abbildung 6.8: bedingte Dichten $P_{\hat{\theta},\hat{\psi}}(\underline{Y} \mid \underline{R} = r)$ $(r = 0{,}1)$ und Dichte der Normalverteilung von $\underline{Y}$ im Selection Modell

Die Dichtefunktion $P_\theta(\underline{Y})$ im Selection Modell weicht erheblich von der Dichtefunktion $P_{\omega,\varepsilon}(\underline{Y})$ im Pattern-Mixture Modell ab. Dies kann auch durch die geschätzten Momente der Zufallsvariable $\underline{Y}$ belegt werden. Im univariaten Pattern-Mixture Modell wird der Erwartungswert $E(\underline{Y}) = \mu$ durch

$$(1-\hat{\varepsilon}_1)\hat{\mu}^{(0)} + \hat{\varepsilon}_1\hat{\mu}^{(1)} \approx 0{,}24\hat{\mu}^{(1)}$$

und die Varianz $Var(\underline{Y}) = \sigma^2$ mittels

$$(1-\hat{\varepsilon}_1)\left(\hat{\sigma}^{(0)}\right)^2 + \hat{\varepsilon}_1\left(\hat{\sigma}^{(1)}\right)^2 + \hat{\varepsilon}_1(1-\hat{\varepsilon}_1)\left(\hat{\mu}^{(1)} - \hat{\mu}^{(0)}\right)^2 \approx 0{,}98 + 0{,}18\left(\hat{\mu}^{(1)}\right)^2$$

[191] Wie in Kapitel 6.3.2.2 gezeigt kann die Schätzung durch den EM-Algorithmus erfolgen.

geschätzt,[192] so dass sich diese Werte von den Schätzern

$$\hat{\mu} = 0{,}2,\ \hat{\sigma}^2 = 1{,}21$$

im Selection Modell im Allgemeinen unterscheiden. Das Ausmaß der Abweichung hängt dabei von dem im Pattern-Mixture Modell festzulegenden Restriktionen bezüglich des Parameters $\mu^{(1)}$ ab.

Im Gegensatz zum Beispiel 6.11 sind die Dichtefunktionen $P_\theta(\underline{Y})$ und $P_{\omega,\varepsilon}(\underline{Y})$ des Selection bzw. Pattern-Mixture Modells identisch, wenn der Ausfallmechanismus vom Typ MCAR ist. Für das Selection Modell gilt unter dieser Voraussetzung[193]

$$\begin{aligned} P_{\theta,\psi}(\mathbf{y}_{obs}, \mathbf{r}) &= \int P_\psi(\mathbf{r} \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) P_\theta(\mathbf{y}_{obs}, \mathbf{y}_{mis}) d\mathbf{y}_{mis} \\ &= \int P_\psi(\mathbf{r}) P_\theta(\mathbf{y}_{obs}, \mathbf{y}_{mis}) d\mathbf{y}_{mis} \\ &= P_\psi(\mathbf{r}) P_\theta(\mathbf{y}_{obs}) \end{aligned} \tag{6.129}$$

und $P_{\omega,\varepsilon}(\mathbf{y}_{obs}, \mathbf{r})$ vereinfacht sich im Pattern-Mixture Modell zu[194]

$$\begin{aligned} P_{\omega,\varepsilon}(\mathbf{y}_{obs}, \mathbf{r}) &= \int P_\varepsilon(\mathbf{r}) P_\omega(\mathbf{y}_{obs}, \mathbf{y}_{mis} \mid \mathbf{r}) d\mathbf{y}_{mis} \\ &= P_\varepsilon(\mathbf{r}) \int P_\omega(\mathbf{y}_{obs}, \mathbf{y}_{mis}) d\mathbf{y}_{mis} \\ &= P_\varepsilon(\mathbf{r}) P_\omega(\mathbf{y}_{obs}). \end{aligned} \tag{6.130}$$

Aus den beiden Verteilungen in (6.129) und (6.130) ist ersichtlich, dass aufgrund der Unabhängigkeit der Indikatormatrix **R** von der Datenmatrix **y** die Parameter von den beiden Modellen identisch sind:

$$\psi = \varepsilon \quad , \quad \theta = \omega \tag{6.131}$$

Diese Identität der Parameter gilt nicht, falls die Werte von $\underline{Y}$ systematisch fehlen. Ist z.B. die MAR-Annahme erfüllt, so lässt sich $P_{\omega,\varepsilon}(\mathbf{y}_{obs}, \mathbf{r})$ wie folgt in den beiden Modellen darstellen:

---

[192] Vgl. Little/Rubin (2002), S. 327.
[193] Vgl. Kastner, C. (2001), S. 58f.
[194] Vgl. Kastner, C. (2001), S.58f.

$$P_{\theta,\psi}(\mathbf{y}_{obs},\mathbf{r}) = \int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs}) P_{\theta}(\mathbf{y}_{obs},\mathbf{y}_{mis}) d\mathbf{y}_{mis}$$
$$= P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs}) P_{\theta}(\mathbf{y}_{obs}) \tag{6.132}$$

$$P_{\omega,\varepsilon}(\mathbf{y}_{obs},\mathbf{r}) = \int P_{\varepsilon}(\mathbf{r}) P_{\omega}(\mathbf{y}_{obs},\mathbf{y}_{mis} \mid \mathbf{r}) d\mathbf{y}_{mis}$$
$$= P_{\varepsilon}(\mathbf{r}) \int P_{\omega}(\mathbf{y}_{obs},\mathbf{y}_{mis} \mid \mathbf{r}) d\mathbf{y}_{mis}$$
$$= P_{\varepsilon}(\mathbf{r}) P_{\omega}(\mathbf{y}_{obs} \mid \mathbf{r}) \tag{6.133}$$

Der Parametervektor ω kann im Pattern-Mixture Modell nicht ohne weitere Annahmen bestimmt werden, obwohl in (6.133) die MAR-Eigenschaft der Daten vorausgesetzt wird. Diese Problematik ist dem Modell inhärent, da sich der Vektor $\omega = (\omega_0,...,\omega_s)$ auf die Parameter der einzelnen Pattern bezieht und deren Schätzwerte nicht allein aus den beobachteten Daten $\mathbf{y}_{obs}$ ermittelt werden können. Im Gegensatz zur MCAR Annahme kann somit die MAR-Eigenschaft nicht ohne weiteres bei Pattern-Mixture Modellen formalisiert werden. Molenberghs et al. (1998) haben gezeigt, dass zumindest bei monotonen Ausfallmustern bestimmte, im Sinne von Little (1993) definierte Identifying Restrictions mit der MAR-Annahme äquivalent ist. Hierfür ist eine Zufallsvariable *M* zu definieren, welche die Anzahl der fehlenden Werte des jeweiligen Pattern beschreibt. Die folgende Darstellung verdeutlicht anhand eines Beispiels mit 5 Zufallsvariablen ($k = 5$) die Festlegung von *M* für ein monotones Ausfallmuster.

i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}
1					
2					
3					?
4				?	?
5				?	?
6		?	?	?	?

R_{i1}	R_{i2}	R_{i3}	R_{i4}	R_{i5}
0	0	0	0	0
0	0	0	0	0
0	0	0	0	1
0	0	0	1	1
0	0	0	1	1
0	1	1	1	1

M
0
0
1
2
2
4

Abbildung 6.9: Definition der Zufallsvariable *M* bei einem monotonen Ausfallmuster

Die bereits erwähnten Restriktionen im Pattern-Mixture Modell, welche für monotone Ausfallmuster identisch mit der MAR-Annahme im Selection Modell sind,[195]

[195] Der Beweis erfolgt in Molenberghs et al. (1998), S. 158.

werden von Molenberghs et al. (1998) als „Available Case Missing Value (ACMV) Restrictions“ bezeichnet. Formell können diese Bedingungen durch

$$P(Y_j \mid Y_1,..,Y_{j-1}, M = m) = P(Y_j \mid Y_1,..,Y_{j-1}, M \le k - j) \quad \forall j > k - m, \forall m > 0 \tag{6.134}$$

dargestellt werden. Die ACMV-Restriktionen implizieren, dass die bedingte Verteilung einer unbeobachteten Zufallsvariable Y_j in einem Pattern m, gegeben die (beobachteten und unbeobachteten) Variablen $Y_1,..Y_{j-1}$, gleich der entsprechenden bedingten Verteilung in einem zusammengefassten Pattern ist, in dem alle Werte von $Y_1,\ldots,Y_j$ beobachtet wurden.

Beispiel 6.12:

Es liege das monotone Ausfallmuster aus Abbildung 6.9 für 5 Zufallsvariablen $Y_1,...,Y_5$ vor. Dann führen die ACMV-Restriktionen in (6.134) zur Äquivalenz der folgenden bedingten Verteilungen:

$$P(Y_5 \mid Y_1,...,Y_4, M = m) = P(Y_5 \mid Y_1,...,Y_4, M = 0) \qquad \forall m > 0$$

$$P(Y_4 \mid Y_1, Y_2, Y_3, M = m) = P(Y_4 \mid Y_1, Y_2, Y_3, M \le 1) \qquad \forall m > 1$$

$$P(Y_3 \mid Y_1, Y_2, M = 4) = P(Y_3 \mid Y_1, Y_2, M \le 2)$$

$$P(Y_2 \mid Y_1, M = 4) = P(Y_2 \mid Y_1, M \le 3)$$

Die folgende Abbildung verdeutlicht am Beispiel der unbekannten bedingten Verteilungen $P(Y_4 \mid Y_1, Y_2, Y_3, M = m)$ für alle $m > 1$, dass zu deren Bestimmung die Daten aus den Pattern $M = 0$ und $M = 1$ gemeinsam verwendet werden.

i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	M
1						**0**
2						**0**
3					?	**1**
4				?	?	2
5				?	?	2
6		?	?	?	?	4

Abbildung 6.10: Darstellung der relevanten Daten zur Bestimmung von

$$P(Y_4 \mid Y_1, Y_2, Y_3, M = m) \text{ für alle } m > 1$$

Die ACMV-Restriktionen sind von den „Complete Case Missing Value (CCMV) Restrictions“, die von Little (1993) eingeführt wurden, zu unterscheiden. Bei letzteren wird vorausgesetzt, dass die bedingte Verteilung einer unbeobachteten Zufallsvariable Y_j in einem Pattern m, gegeben die Variablen $Y_1,\ldots,Y_{j-1}$, gleich der entsprechenden bedingten Verteilung in dem Pattern der vollständig beobachteten Daten ($M = 0$) ist:[196]

$$P(Y_j \mid Y_1,..,Y_{j-1}, M = m) = P(Y_j \mid Y_1,..,Y_{j-1}, M = 0) \quad \forall j > k - m, \forall m > 0 \qquad (6.135)$$

Werden lediglich zwei Zufallsvariablen Y_1 und Y_2 betrachtet, von denen eine Variable stets beobachtet wird (univariater Datenausfall), so stimmen die ACMV-Restriktionen und CCMV-Restriktionen überein. In Kapitel 6.4.2.2 wurde nachgewiesen, dass die ACMV- bzw. CCMV-Restriktion (6.68) in dem betrachteten Modell identisch mit der MAR-Annahme ist, da lediglich die Zufallsvariable Y_2 fehlende Werte aufweist und somit ein monotones Ausfallmuster vorliegt. Ist das Ausfallmuster jedoch nicht monoton, gilt die Äquivalenz von ACMV-Restriktionen und MAR-Bedingung nicht, wie das folgende Beispiel zeigt.

Beispiel 6.13:

Es sei $\underline{Y} = (Y_1, Y_2)$ eine zweidimensionale Zufallsvariable, deren Beobachtung von Y_1 und Y_2 durch die bivariate Indikatorvariable $\underline{R} = (R_1, R_2)$ beschrieben wird. Der Responsemechanismus bezüglich eines nicht-monotonen Ausfallmusters sei wie folgt definiert:[197]

$$P(\underline{R} = (1,1) \mid \underline{y}) = p$$

$$P(R = (0,1) \mid \underline{y}) = f(y_1)$$

$$P(R = (1,0) \mid \underline{y}) = f(y_2)$$

$$P(R = (0,0) \mid \underline{y}) = 1 - p - f(y_1) - f(y_2) \qquad (6.136)$$

[196] Vgl. Little (1993), S. 128.
[197] Vgl. Molenberghs et al. (1998), S. 156.

Die MAR-Annahme ist in diesem Beispiel erfüllt, da das Fehlen von Y_1 allein von Y_2 – angegeben durch den funktionalen Zusammenhang $f(y_2)$ – abhängig ist und die Nichtbeobachtung von Y_2 durch $f(y_1)$ beschrieben wird.

Wird vorausgesetzt, dass die ACMV-Restriktionen im Pattern-Mixture Modell gelten, sind u.a. die bedingten Wahrscheinlichkeiten

$$P\left(y_1 \mid y_2, \underline{R} = (1,0)\right) = P\left(y_1 \mid y_2, \underline{R} = (0,0)\right) \tag{6.137}$$

und

$$P\left(y_2 \mid y_1, \underline{R} = (0,1)\right) = P\left(y_2 \mid y_1, \underline{R} = (0,0)\right) \tag{6.138}$$

identisch.

Für die gemeinsame Wahrscheinlichkeit von y_1 und y_2 in den Pattern $\underline{R} = (1,0)$ und $\underline{R} = (0,1)$ gilt somit:[198]

$$P\left(y_1, y_2 \mid \underline{R} = (1,0)\right) = P\left(y_2 \mid \underline{R} = (1,0)\right) P\left(y_1 \mid y_2, \underline{R} = (0,0)\right) \tag{6.139}$$

$$P\left(y_1, y_2 \mid \underline{R} = (0,1)\right) = P\left(y_1 \mid \underline{R} = (0,1)\right) P\left(y_2 \mid y_1, \underline{R} = (0,0)\right) \tag{6.140}$$

Aus (6.140) folgt durch Anwendung des Satzes von Bayes sowie durch Faktorisierung[199]

$$\frac{P\left(\underline{R} = (0,1) \mid y_1, y_2\right) P\left(y_1, y_2\right)}{P\left(\underline{R} = (0,1)\right)} = \frac{P\left(\underline{R} = (0,1) \mid y_1\right) P\left(y_1\right)}{P\left(\underline{R} = (0,1)\right)} \frac{P\left(\underline{R} = (0,0) \mid y_1, y_2\right) P\left(y_1, y_2\right)}{P\left(\underline{R} = (0,0) \mid y_1\right) P\left(y_1\right)}$$

und somit

$$P\left(\underline{R} = (0,1) \mid y_1, y_2\right) = P\left(\underline{R} = (0,1) \mid y_1\right) \frac{P\left(\underline{R} = (0,0) \mid y_1, y_2\right)}{P\left(\underline{R} = (0,0) \mid y_1\right)}. \tag{6.141}$$

Weiterhin ist nach (6.136) der Datenausfall von Y_2 nicht von dem Wert der Zufallsvariable Y_2 abhängig

$$P\left(\underline{R} = (0,1) \mid y_1, y_2\right) = P\left(\underline{R} = (0,1) \mid y_1\right) = f(y_1),$$

und demzufolge wird die Wahrscheinlichkeit für die Beobachtung beider Zufallsva-

[198] Vgl. Molenberghs et al. (1998), S. 157.
[199] Vgl. Molenberghs et al. (1998), S. 157.

riablen nicht von y_2 beeinflusst:[200]

$$P(\underline{R} = (0,0) \mid y_1, y_2) = P(\underline{R} = (0,0) \mid y_1)$$

In gleicher Weise kann anhand der Bedingung (6.139) gezeigt werden, dass $P(\underline{R} = (0,0) \mid y_1, y_2)$ auch nicht von y_1 abhängig ist. Somit ist bei Einhaltung der ACMV-Restriktionen nachgewiesen, dass

$$P(\underline{R} = (0,0) \mid y_1, y_2) = 1 - p - f(y_1) - f(y_2) = 1 - p - c_1 - c_2 \tag{6.142}$$

mit den zwei Konstanten c_1 und c_2 gilt. Der Ausfallmechanismus in (6.136) hängt nicht mehr von den Werten von Y_1 und Y_2 ab und die Daten fehlen zufällig im Sinne von MCAR. Damit sind bei diesem nicht-monotonen Ausfallmuster die ACMV-Restriktionen *nicht* mit der MAR-Annahme vereinbar.[201]

Als wesentliches Ergebnis ist festzuhalten, dass bei systematischem Datenausfall (bzw. bei Verletzung der MCAR-Annahme) Pattern-Mixture Modelle und Selection Modelle im Allgemeinen zu unterschiedlichen Ergebnissen führen.[202] Die korrekte Ersetzung von fehlenden Werten unter dem Datenausfall vom Typ MNAR erfordert bei den Pattern-Mixture Modellen die Gültigkeit der Identifying Restrictions und der Verteilungsannahmen bezüglich $\underline{Y}$ gegeben $\underline{R} = \underline{r}$, während bei Anwendung der Selection Modelle eine Spezifizierung der Verteilungstypen von dem Ausfallmechanismus sowie von $\underline{Y}$ notwendig ist. Da diese Voraussetzungen bei beiden Ansätzen nicht überprüfbar sind, ist es nahe liegend, unterschiedliche Identifying Restrictions bzw. Verteilungstypen anzunehmen und eine Sensitivitätsanalyse durchzuführen. Nach der multiplen Analyse der vervollständigten Daten kann ein Intervall für die Schätzung eines Parameters angegeben werden, welches den unterschiedlichen (Verteilungs-)annahmen bzw. Restriktionen während des Ersetzungsprozesses Rechnung trägt. Die Sensitivitätsanalyse ist innerhalb eines Pattern-Mixture Modells einfacher zu realisieren, da in diesem Ansatz Zusammenhänge zwischen den Momenten der Zufallsvariable $\underline{Y}$ und den nicht identifizierbaren Parametern hergestellt werden können. Ein weiterer Vorteil der Pattern-Mixture Modelle besteht in der Möglichkeit,

[200] Vgl. Formel (6.141).
[201] Vgl. Molenberghs (1998), S. 157.
[202] Vgl. Little (1994), S. 472.

zumindest die Verteilungsannahmen im Pattern der vollständig beobachteten Merkmalsträger ($r = 0$) mit den gängigen statistischen Testverfahren (Kolmogorov-Smirnov Anpassungstest, χ^2-Anpassungstest)[203] überprüfen zu können. Diese Tests können in Selection Modellen aufgrund der bereits im univariaten, stetigen Fall kompliziert zu bestimmenden Dichtefunktion $P_{\theta,\psi}(\underline{Y} \mid \underline{R} = 0)$ häufig nicht durchgeführt werden.

Eine der Pattern-Mixture Modellierung immanente Problematik ist die Interpretierbarkeit der Identifying Restrictions, die sich in vielen Fällen als schwierig erweist. Dieser Nachteil wird zumindest für monotone Ausfallmuster durch die ACMV-Restriktionen abgeschwächt, da letztere eine Berücksichtigung der (leicht zu interpretierenden) MAR-Annahme in den Pattern-Mixture Modellen ermöglichen. Insofern können die ACMV-Restriktionen auch als Ausgangspunkt der Sensitivitätsanalyse dienen, und durch gezielte Variierung einzelner Restriktionen können die Auswirkungen auf die Schätzungen untersucht werden. Pattern-Mixture Modelle stellen somit einen flexiblen Ansatz dar, um fehlende Werte unter nicht ignorierbaren Ausfallmechanismen zu behandeln.

[203] Vgl. Hartung (1999), S. 182ff.

7 Simulationsstudien zur Bewertung der Behandlungsverfahren

7.1 Ziele der Simulationen

In den vorangegangenen Kapiteln 5 und 6 wurden Verfahren diskutiert, die bei ignorierbaren bzw. nicht ignorierbaren Ausfallmechanismen unverzerrte Parameterschätzungen ermöglichen. Wie bereits in Kapitel 5.3 dargestellt, finden insbesondere likelihood-basierten Verfahren, welche die Gültigkeit der MAR-Annahme voraussetzen, verbreitet Anwendung bei der Behandlung von fehlenden Werten. Unter diesem Aspekt stellt sich die Frage, wie sich eine Anwendung dieser Verfahren auf die Schätzungen auswirkt, falls die zugrunde liegende MAR-Annahme verletzt ist. Diese Problematik wurde bereits in den Kapiteln 6.2.3 und 6.4.2.4 anhand von Beispielen diskutiert,[204] und im Folgenden wird diese Betrachtung erweitert, indem unter verschiedenen Szenarien, die sich durch eine Variierung der Ausfallquote sowie des Ausfallmechanismus ergeben, Simulationen zur Feststellung des Bias der MAR-basierten Methoden durchgeführt werden.

7.2 Untersuchung bei diskreten Variablen

7.2.1 Allgemeiner Simulationsaufbau

Bereits in Kapitel 3 wurde auf die vorherrschende Meinung in der statistischen Literatur verwiesen, die eine Behandlung von fehlenden Werten mittels Verfahren, welche die Erfüllung der MAR-Annahme voraussetzen, in umfangreichen Datenbeständen postuliert. Bei zahlreichen statistischen Analysen werden jedoch nur wenige Variablen betrachtet, und eine Ersetzung mittels MAR-basierter Verfahren erscheint unter diesen Umständen zweifelhaft.

Um insbesondere diese Fälle genauer zu untersuchen, sei das in Kapitel 6.2.3 verwendete Modell für diskrete Zufallsvariablen der Ausgangspunkt der folgenden Simulationen. In diesem Modell wurden zwei binäre Variablen Y_1 und Y_2 betrachtet, wobei Y_1 vollständig in der Stichprobe beobachtet wurde und Y_2 fehlende Werte aufweist. Weiterhin wurde die Gültigkeit der bedingten Unabhängigkeitsbeziehung

[204] Vgl. Beispiel 6.3 und Beispiel 6.9.

$$R_2 \perp\!\!\!\perp Y_1 \mid Y_2 \tag{7.1}$$

vorausgesetzt, so dass für die bedingte Verteilung $P(R_2 \mid y_1, y_2)$ gilt:

$$P(R_2 \mid y_1, y_2) = P(R_2 \mid y_2) \tag{7.2}$$

In der Simulationsstudie ist es somit ausreichend, die bedingte Verteilung $P(R_2 \mid y_2)$ zu spezifizieren. Die Ausfallwahrscheinlichkeit hängt dabei von den partiell nicht beobachteten Ausprägungen der Zufallsvariable Y_2 ab, so dass der Ausfallmechanismus vom Typ MNAR ist:

$$P(\mathbf{R} \mid \mathbf{y}) = \prod_{i=1}^{n} P(R_{i2} \mid y_{i1}, y_{i2}) = \prod_{i=1}^{n} P(R_{i2} \mid y_{i2}) = \prod_{i=1}^{n} P(R_{i2} \mid \underline{y}_{mis,i}, \underline{y}_{obs,i}) \neq P(\mathbf{R} \mid \mathbf{y}_{obs})$$

Um die MAR- und MNAR-basierten Verfahren bezüglich des Bias von Parameterschätzwerten zu beurteilen, sei die folgende Verteilung von Y_1 und Y_2 der Ausgangspunkt der Betrachtungen:

	$Y_2 = 0$	$Y_2 = 1$	
$Y_1 = 0$	0,2	0,3	*0,5*
$Y_1 = 1$	0,4	0,1	*0,5*
	0,6	*0,4*	*1*

Tabelle 7.1: Verteilung der Zufallsvariablen Y_1 und Y_2

Weiterhin sei die (marginale) Ausfallwahrscheinlichkeit durch

$$P(R_2 = 1) = 0{,}4$$

vorgegeben. Unter Kenntnis dieser Wahrscheinlichkeiten können die bedingten Verteilungen $P(R_2 \mid Y_2 = b)$ $(b = 0,1)$ in Abhängigkeit von einem Parameter λ_1 $(-0{,}4 \le \lambda_1 \le 0{,}2\overline{6})$ beschrieben werden:

$$P(R_2 = 1 \mid Y_2 = 0) = P(R_2 = 1) + \lambda_1 = 0{,}4 + \lambda_1$$

$$P(R_2 = 0 \mid Y_2 = 0) = 1 - P(R_2 = 1 \mid Y_2 = 0) = 0{,}6 - \lambda_1$$

$$\begin{aligned} P(R_2 = 1 \mid Y_2 = 1) &= \frac{P(R_2 = 1, Y_2 = 1)}{P(Y_2 = 1)} = \frac{P(R_2 = 1) - P(R_2 = 1, Y_2 = 0)}{P(Y_2 = 1)} \\ &= \frac{P(R_2 = 1) - P(R_2 = 1 \mid Y_2 = 0) P(Y_2 = 0)}{P(Y_2 = 1)} \\ &= \frac{0{,}4 - \left[(0{,}4 + \lambda_1) 0{,}6\right]}{0{,}4} \\ &= 0{,}4 - 1{,}5\,\lambda_1 \end{aligned}$$

$$P(R_2 = 0 \mid Y_2 = 1) = 1 - P(R_2 = 1 \mid Y_2 = 1) = 0{,}6 + 1{,}5\,\lambda_1 \qquad (7.3)$$

Der Parameter λ_1 ist als Index für die Nicht-Ignorierbarkeit des Ausfallmechanismus zu interpretieren, wobei für $\lambda_1 = 0$ der Ausfallmechanismus sowohl von Y_1 als auch von Y_2 unabhängig ist. Letzteres gilt, da in diesem Fall

$$P(R_2 = 1 \mid Y_2 = 0) = P(R_2 = 1 \mid Y_2 = 1) = 0{,}4$$

bzw.

$$R_2 \perp\!\!\!\perp Y_2$$

ist, und gemeinsam mit der (annahmegemäß geltenden) bedingten Unabhängigkeitsbeziehung

$$R_2 \perp\!\!\!\perp Y_1 \mid Y_2$$

kann die marginale Unabhängigkeit

$$R_2 \perp\!\!\!\perp Y_1$$

abgeleitet werden. Somit gilt für $\lambda_1 = 0$ die MCAR-Annahme, während für $\lambda_1 \neq 0$ der Ausfallmechanismus vom Typ MNAR ist. Die folgende Abbildung zeigt, dass der Einfluss der Zufallsvariable Y_2 auf den Datenausfall mit höheren Werten von $|\lambda_1|$ zunimmt.

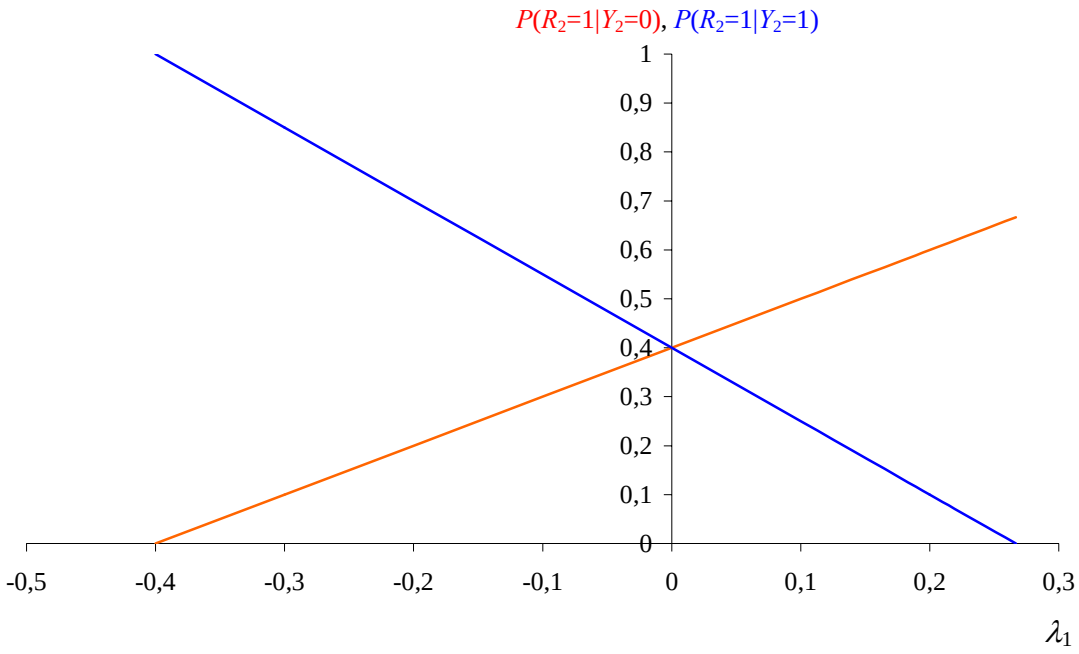


Abbildung 7.1: Bedingte Wahrscheinlichkeiten $P(R_2 = 1 | Y_2 = 0)$ und $P(R_2 = 1 | Y_2 = 1)$ in Abhängigkeit von λ_1

Aus der Verteilung $P(Y_1, Y_2)$ in Tabelle 7.1 und der bedingten Verteilung $P(R_2 \mid Y_2)$ in (7.3) kann die gemeinsame Verteilung $P(R_2, Y_1, Y_2)$ mittels

$$P(R_2, Y_1, Y_2) = P(R_2 \mid Y_1, Y_2)\, P(Y_1, Y_2) = P(R_2 \mid Y_2)\, P(Y_1, Y_2)$$

bestimmt werden:

	$Y_2 = 0$		$Y_2 = 1$		
	$Y_1 = 0$	$Y_1 = 1$	$Y_1 = 0$	$Y_1 = 1$	
$R_2 = 0$	0,12-0,2λ_1	0,24-0,4λ_1	0,18+0,45λ_1	0,06+0,15λ_1	*0,6*
$R_2 = 1$	0,08+0,2λ_1	0,16+0,4λ_1	0,12-0,45λ_1	0,04-0,15λ_1	*0,4*
	0,2	*0,4*	*0,3*	*0,1*	*1*

Tabelle 7.2: Gemeinsame Verteilung der Zufallsvariablen R_2, Y_1 und Y_2 in Abhängigkeit von λ_1

Ausgehend von dieser gemeinsamen Verteilung wird *für jeden untersuchten Wert des Parameters* λ_1 eine Stichprobe vom Umfang $n = 100$ gezogen. In Kapitel 6.2.3 wurde gezeigt, dass bei Gültigkeit von (7.1) die Parameter θ_{ab} ($a = 0,1$; $b = 0,1$) der Verteilung

$$P_{\theta_{ab}}(Y_1 = a, Y_2 = b) = \theta_{ab} \qquad a = 0,1,\ b = 0,1$$

unverzerrt aus den beobachteten Daten der Stichprobe geschätzt werden können.[205] Zur Evaluierung der Behandlungsmethoden wird der Schätzer $\hat{\theta}_{10}$ des (unter diesem Ausfallmechanismus korrekten) MNAR-basierten Verfahrens aus Kapitel 6.2.3 mit demjenigen des EM-Algorithmus, der in seiner allgemeinen Form die Gültigkeit der MAR-Annahme voraussetzt, für jeden Wert des Parameters λ_1 verglichen. Zusätzlich wird ein Eliminierungsverfahren in Form der Complete Case Analysis aus Kapitel 5.1.2 angewendet, um die Auswirkungen der MCAR-Annahme auf die Parameterschätzungen zu verdeutlichen.

Im Rahmen der Simulation wird die beschriebene Vorgehensweise 100 mal wiederholt. Die einzelnen Schätzwerte, die bei der Anwendung eines bestimmten Verfahrens ermittelt wurden, werden durch Bildung des arithmetischen Mittels zu einem Wert zusammengefasst. Weiterhin wird – ausgehend von jeder einzelnen Schätzung eines Verfahrens – ein 95%-Konfidenzintervall für den wahren Parameter ermittelt und überprüft, wie oft dieser in den insgesamt 100 Konfidenzintervallen liegt. Ist dieser Anteil, der auch als Überdeckung bezeichnet wird, deutlich unter 95%, so kann davon ausgegangen werden, das die jeweilige Behandlungsmethode zu verzerrten Ergebnissen führt.[206] Der Bias wird ebenfalls anhand des Standardfehlers des Schätzers, der sich bei vollständiger Beobachtung aller Werte in einer Stichprobe vom Umfang $n = 100$ ergeben würde, beurteilt: Ist die Abweichung des (aus den 100 Stichproben ermittelten) Schätzwerts von dem wahren Wert größer als die Hälfte des Standardfehlers, so ist von einer starken Verzerrung durch die Methode auszugehen.[207]

---

[205] Vgl. Schätzer von θ_{ab} in (6.26)-(6.29).
[206] Vgl. Schafer/Graham (2002), S. 156. Schafer/Graham betrachten in ihrer Simulation einen Anteil von <90% bereits als problematisch.
[207] Vgl. Schafer/Graham (2002), S. 156f.

Innerhalb der Simulation wird ebenfalls untersucht, inwieweit die Ausfallquote

$$\frac{n-q}{n}\cdot 100\%$$ [208]

einen Einfluss auf das Ausmaß der Verzerrung besitzt. Hierfür werden für verschiedene Ausfallquoten zwischen 20% und 70% die Ergebnisse der Behandlungsmethoden unter konstanten Bedingungen (λ_1 = -0,2) miteinander verglichen.

Wie bereits erläutert, ist die unverzerrte Parameterschätzung mittels des MNAR-basierten Verfahrens aus Kapitel 6.2.3 an die Gültigkeit der bedingten Unabhängigkeitsbeziehung (7.1) geknüpft. Unter diesem Gesichtspunkt ist zu untersuchen, wie robust dieses Verfahren bei Verletzung der Annahme ist. Hierzu wird ein weiterer Parameter λ_2 ($-0{,}2 \le \lambda_2 \le 0{,}2$) eingeführt, der die Stärke der Abhängigkeit zwischen R_2 und Y_1 gegeben die Zufallsvariable Y_2 beschreiben soll. Die bedingten Wahrscheinlichkeiten $P(R_2 = 1 \mid Y_1 = a, Y_2 = b)$ (a = 0,1, b = 0,1) seien annahmegemäß in der folgenden Weise von dem Parameter λ_2 abhängig:[209]

$$P(R_2 = 1 \mid Y_1 = 0, Y_2 = 0) = P(R_2 = 1 \mid Y_2 = 0) - \lambda_2 = 0{,}4 + \lambda_1 - \lambda_2 = 0{,}2 - \lambda_2$$

$$P(R_2 = 1 \mid Y_1 = 1, Y_2 = 0) = P(R_2 = 1 \mid Y_2 = 0) + \lambda_2 = 0{,}4 + \lambda_1 + \lambda_2 = 0{,}2 + \lambda_2$$

$$P(R_2 = 1 \mid Y_1 = 0, Y_2 = 1) = P(R_2 = 1 \mid Y_2 = 1) - \lambda_2 = 0{,}4 - 1{,}5\lambda_1 - \lambda_2 = 0{,}7 - \lambda_2$$

$$P(R_2 = 1 \mid Y_1 = 1, Y_2 = 1) = P(R_2 = 1 \mid Y_2 = 1) + \lambda_2 = 0{,}4 - 1{,}5\lambda_1 + \lambda_2 = 0{,}7 + \lambda_2 \qquad (7.4)$$

Im Fall $\lambda_2 = 0$ gilt die bedingte Unabhängigkeitsbeziehung

$R_2 \perp\!\!\!\perp Y_1 \mid Y_2$,

während für $\lambda_2 \neq 0$ die Zufallsvariable Y_1 den Datenausfall in einem von λ_2 abhängigen Ausmaß beeinflusst:

---

[208] q bezeichnet dabei die Anzahl der vollständig beobachteten Untersuchungseinheiten.
[209] Es gelten die Simulationsbedingungen λ_1 = -0,2 , 40% Ausfallquote.

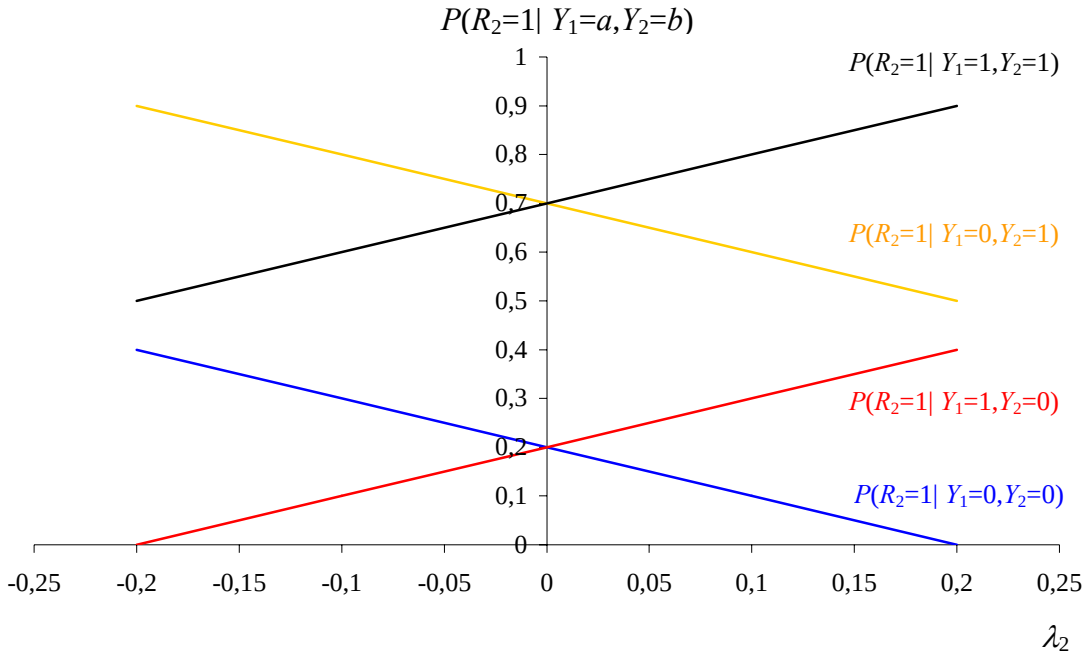


Abbildung 7.2: Bedingte Wahrscheinlichkeiten in Abhängigkeit von λ_2

Ausgehend von den bedingten Wahrscheinlichkeiten $P(R_2 = 1 | Y_1 = a, Y_2 = b)$ (a =0,1, b = 0,1) und der Verteilung $P(Y_1, Y_2)$ in Tabelle 7.1 kann die gemeinsame Verteilung $P(R_2, Y_1, Y_2)$ bestimmt werden:

	$Y_2 = 0$		$Y_2 = 1$		
	$Y_1 = 0$	$Y_1 = 1$	$Y_1 = 0$	$Y_1 = 1$	
$R_2 = 0$	0,16+0,2λ_2	0,32-0,4λ_2	0,09+0,3λ_2	0,03-0,1λ_2	*0,6*
$R_2 = 1$	0,04-0,2λ_2	0,08+0,4λ_2	0,21-0,3λ_2	0,07+0,1λ_2	*0,4*
	0,2	*0,4*	*0,3*	*0,1*	*1*

Tabelle 7.3: Gemeinsame Verteilung der Zufallsvariablen R_2, Y_1 und Y_2 in Abhängigkeit von λ_2 (λ_1 = -0,2)

Durch Variierung des Parameters λ_2 können Abweichungen von der bedingten Unabhängigkeitsannahme analysiert und Schlussfolgerungen bezüglich der Anwendbarkeit von MNAR-basierten Methoden gezogen werden. Abschließend sind diese Ergebnisse mit den Schätzern unter der MAR- bzw. MCAR-Annahme zu verglei-

chen, deren Bestimmung durch den EM-Algorithmus bzw. die Complete Case Analysis erfolgt.

7.2.2 Durchführung und Ergebnisse der Simulation

Im letzten Kapitel wurde der Parameter λ_1 eingeführt, durch den der Zusammenhang zwischen dem Datenausfall und der Zufallsvariable Y_2 unter der Prämisse, dass

$R_2 \perp\!\!\!\perp Y_1 \mid Y_2$

gilt, beschrieben wird. Die Datensimulationen erfolgen anhand verschiedener Werte von λ_1, so dass die Auswirkungen der einzelnen Behandlungsverfahren auf die Schätzung des Parameters

$\theta_{10} = 0{,}4$

bei unterschiedlicher Abhängigkeit von R_2 und Y_2 erfasst werden können. Die folgende Abbildung zeigt für jede der drei Behandlungsmethoden den durchschnittlichen Schätzer $\hat{\theta}_{10}$ aus 100 Stichproben, wenn der Parameter λ_1 zwischen -0,4 und 0,26 variiert wurde und die Ausfallquote konstant bei 40% liegt.

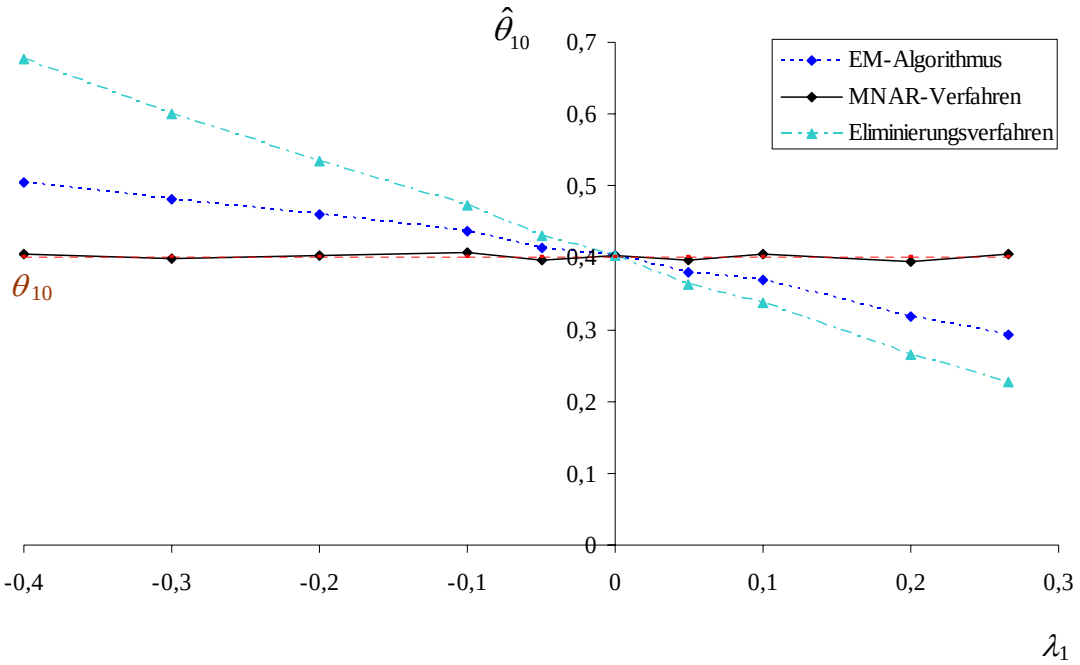


Abbildung 7.3: Schätzung des Parameters θ_{10} unter Variierung von λ_1

Der Parameterschätzer $\hat{\theta}_{10}$ ist bei Anwendung des MNAR-Verfahrens unter allen untersuchten Zusammenhängen zwischen R und Y_2 unverzerrt.[210] Der EM-Algorithmus führt lediglich bei einem geringem Zusammenhang ($-0{,}05 \leq \lambda_1 \leq 0{,}05$) zu unverzerrten Ergebnissen, während bei der Anwendung des Eliminierungsverfahrens lediglich im Fall von MCAR ($\lambda_1 = 0$) kein Bias zu verzeichnen ist. Die Verzerrung durch die Behandlungsmethoden kann weiterhin durch den Anteil der (aus den 100 Stichproben resultierenden) Konfidenzintervalle für θ_{10}, in denen der wahre Parameter liegt, beurteilt werden. Das folgende Diagramm zeigt diese Überdeckung für das MNAR-basierte Verfahren sowie den EM-Algorithmus unter verschiedenen Werten des Parameters λ_1.[211]

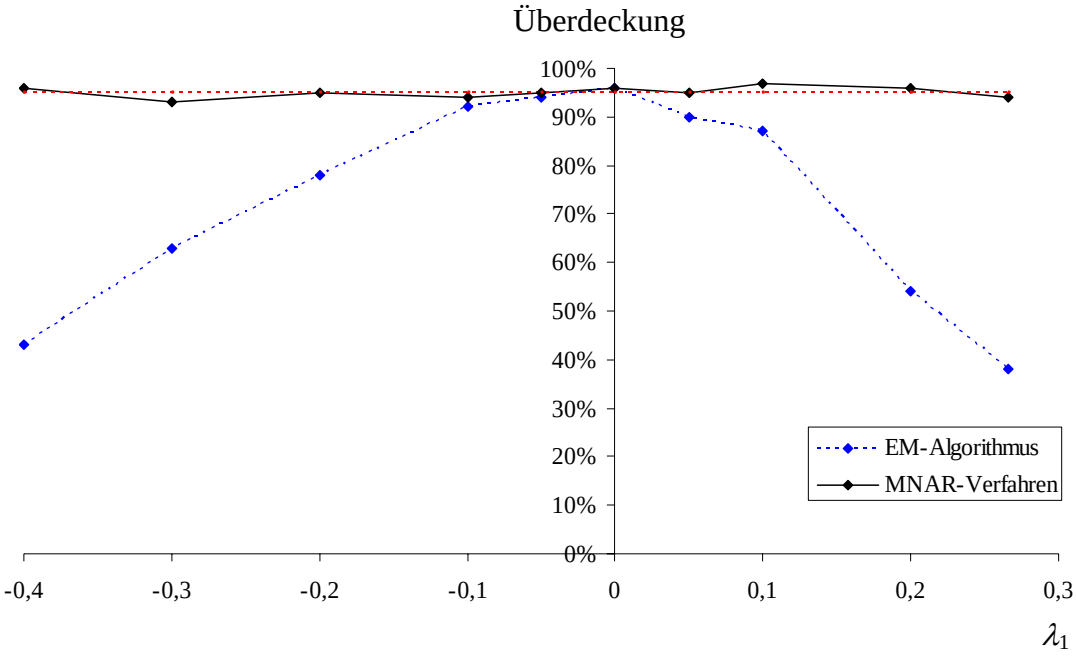


Abbildung 7.4: Überdeckung unter Variierung von λ_1

Da die Überdeckung bei Anwendung des MNAR-basierten Verfahrens stets über 90% liegt, ist in diesem Fall von einer unverzerrten Schätzung des Parameters θ_{10} auszugehen. Hingegen sinkt die Überdeckung beim EM-Algorithmus bereits bei rela-

[210] Vgl. Anhang A.4.

[211] Das Eliminierungsverfahren wurde aufgrund der geringeren effektiven Stichprobengröße nicht berücksichtigt, da die Intervalle eine wesentlich höhere Breite aufweisen und die Aussagekraft beeinträchtigt wird.

tiv niedrigen Werten von $|\lambda_1|$ ($\lambda_1 < -0{,}1$ bzw. $\lambda_1 > 0{,}05$) unter 90%. Der EM-Algorithmus erweist sich somit als nicht robust gegenüber Abweichungen von der MAR-Annahme.

Aus Abbildung 7.3 ist ersichtlich, dass der Bias des Parameterschätzers $\hat{\theta}_{10}$ bei Anwendung des Eliminierungsverfahrens stets größer ist als bei der Behandlung durch den EM-Algorithmus. Dies gilt jedoch nicht für alle Parameterschätzer $\hat{\theta}_{ab}$ ($a = 0{,}1$, $b = 0{,}1$), wie die folgende Abbildung am Beispiel von $\hat{\theta}_{00}$ zeigt.

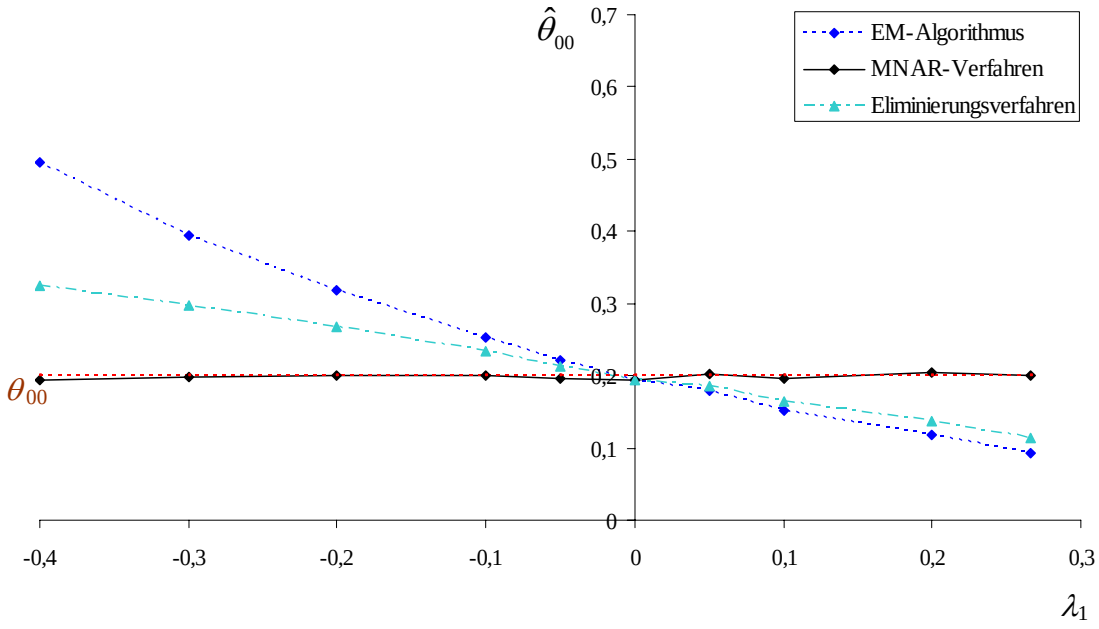


Abbildung 7.5: Schätzung des Parameters θ_{00} unter Variierung von λ_1

Um die Verzerrung *aller* Schätzer zu beurteilen, ist zu überprüfen, ob die mittels der jeweiligen Behandlungsmethode geschätzten absoluten Häufigkeiten

$$\hat{n}_{ab} = n\hat{\theta}_{ab} \qquad a = 0{,}1\,,\, b = 0{,}1$$ [212]

Realisationen der multinomialverteilten Zufallsvariable

$$(N_{00}, N_{10}, N_{01}, N_{11}) \sim M(n, \theta_{00}, \theta_{10}, \theta_{01}, \theta_{11})$$

[212] $\hat{n}_{ab}$ ist somit die geschätzte Anzahl der Merkmalsträger mit $Y_1 = a$ und $Y_2 = b$ in der Stichprobe.

sind. Hierzu wurde für jede Behandlungsmethode ein χ^2-Anpassungstest mit der Testfunktion

$$g(\underline{y}_{obs,1},\ldots,\underline{y}_{obs,n}) = \sum_{a=0}^{1}\sum_{b=0}^{1}\frac{(\hat{n}_{ab} - n\theta_{ab})^2}{n\theta_{ab}} = \sum_{a=1}^{2}\sum_{b=1}^{2}\frac{n(\hat{\theta}_{ab} - \theta_{ab})^2}{\theta_{ab}}$$

durchgeführt, und die in Anhang A.5 aufgeführten p-Werte der Tests zeigen, dass der EM-Algorithmus unter nahezu allen Bedingungen ($-0{,}2 \le \lambda_1 \le 0{,}266$) bessere Ergebnisse als das Eliminierungsverfahren liefert. Somit wird die theoretische Erkenntnis aus Kapitel 5, dass die MAR-basierten Methoden den Eliminierungsverfahren im Rahmen der Behandlung von fehlenden Werten vorzuziehen sind, durch diese Simulationsstudie bestätigt.

In der bisherigen Simulation wurde die Ausfallquote konstant bei 40% belassen. Im Folgenden soll diese Quote unter sonst konstanten Bedingungen ($\lambda_1 = -0{,}2$) variiert werden, um deren Einfluss auf die Schätzung des Parameters θ_{10} beurteilen zu können. Die Ergebnisse, die sich bei Anwendung der einzelnen Behandlungsmethoden ergeben, sind in den folgenden beiden Abbildungen zusammengefasst.

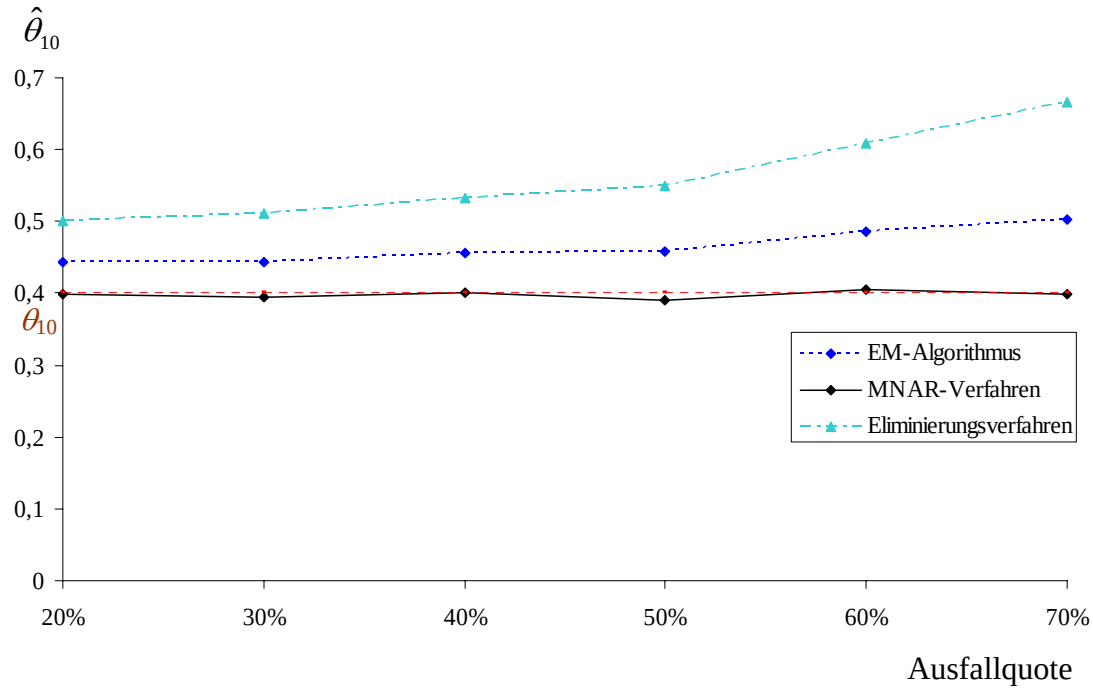


Abbildung 7.6: Schätzung des Parameters θ_{10} bei verschiedenen Ausfallquoten

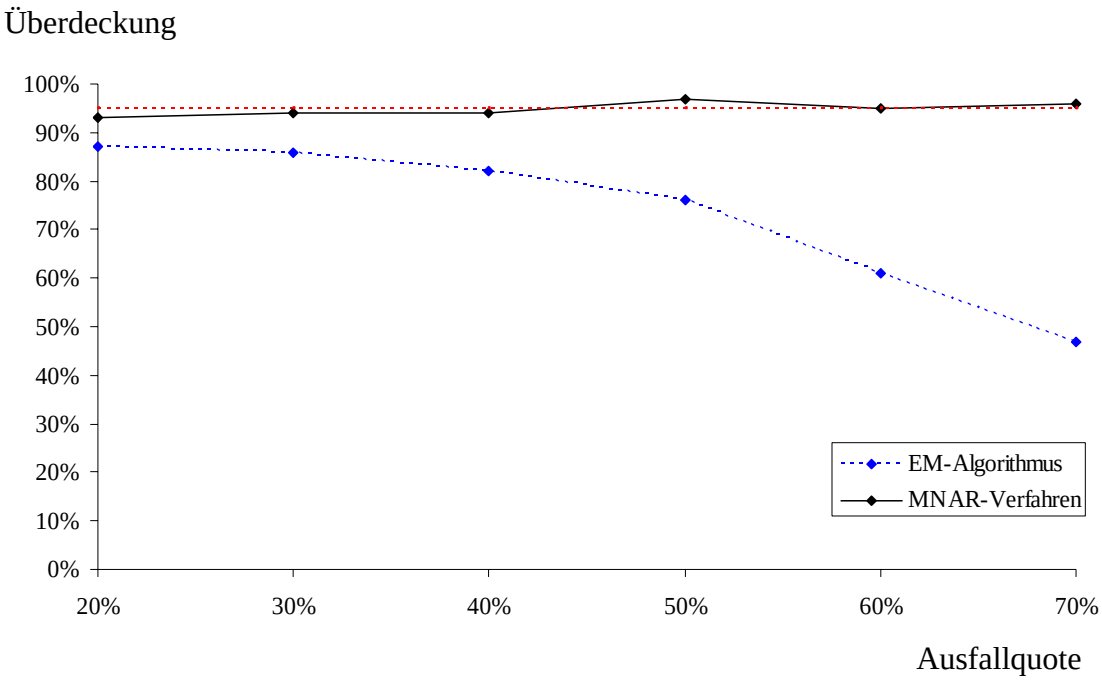


Abbildung 7.7: Überdeckung bei verschiedenen Ausfallquoten

Der Vergleich der Schätzwerte bei einer hohen Ausfallquote verdeutlicht erneut, dass lediglich bei der MNAR-basierten Methode der Parameter θ_{10} unverzerrt geschätzt wird. Diese Feststellung wird durch die Auswertung der Überdeckung bestätigt. Hingegen ist die Anwendung von MCAR- bzw. MAR-basierten Verfahren nicht gerechtfertigt, da unter allen untersuchten Bedingungen die Parameterschätzer verzerrt sind.[213] Weiterhin kann festgestellt werden, dass das Eliminierungsverfahren unter allen Simulationsbedingungen zu größeren Abweichungen der Schätzwerte vom wahren Parameter führt als der EM-Algorithmus. Dies steht im Einklang mit den Ergebnissen der χ^2-Anpassungstests im Anhang A.7, die auf eine insgesamt geringere Verzerrung aller Parameterschätzer $\hat{\theta}_{ab}$ (a = 0,1, b = 0,1) hinweisen. Insbesondere kann bei einer Ausfallquote von 20% die Nullhypothese, dass die auf Basis des EM-Algorithmus geschätzten absoluten Häufigkeiten Realisationen der multinomialverteilten Zufallsvariable

$$(N_{00}, N_{10}, N_{01}, N_{11}) \sim M(n, \theta_{00}, \theta_{10}, \theta_{01}, \theta_{11})$$

[213] Vgl. Anhang A.6.

sind, zu einem Signifikanzniveau von 5% nicht abgelehnt werden. Bei einer Behandlung durch das Eliminierungsverfahren ist die Nullhypothese unter allen untersuchten Ausfallquoten zu verwerfen.

Die bisherigen Untersuchungen beschränkten sich auf simulierte Daten, für welche die bedingte Unabhängigkeitsbeziehung

$R_2 \perp\!\!\!\perp Y_1 \mid Y_2$

erfüllt ist. Es wurde gezeigt, dass die MNAR-basierte Methode unter allen bisher untersuchten Simulationsbedingungen unverzerrte Schätzer liefert. Im Weiteren soll die Robustheit des MNAR-basierten Verfahrens analysiert werden, indem unter der Vorgabe eines Zusammenhangs von R_2 und Y_1 gegeben Y_2 die Datensimulation erfolgt. Die Stärke des Zusammenhangs wird dabei durch den Wert des Parameters λ_2 bestimmt.[214] In den folgenden Abbildungen werden die wesentlichen Simulationsergebnisse, die aus der Variierung von λ_2 unter sonst gleichen Bedingungen (40% Ausfallquote, λ_1 = -0,2) resultieren, verdeutlicht.

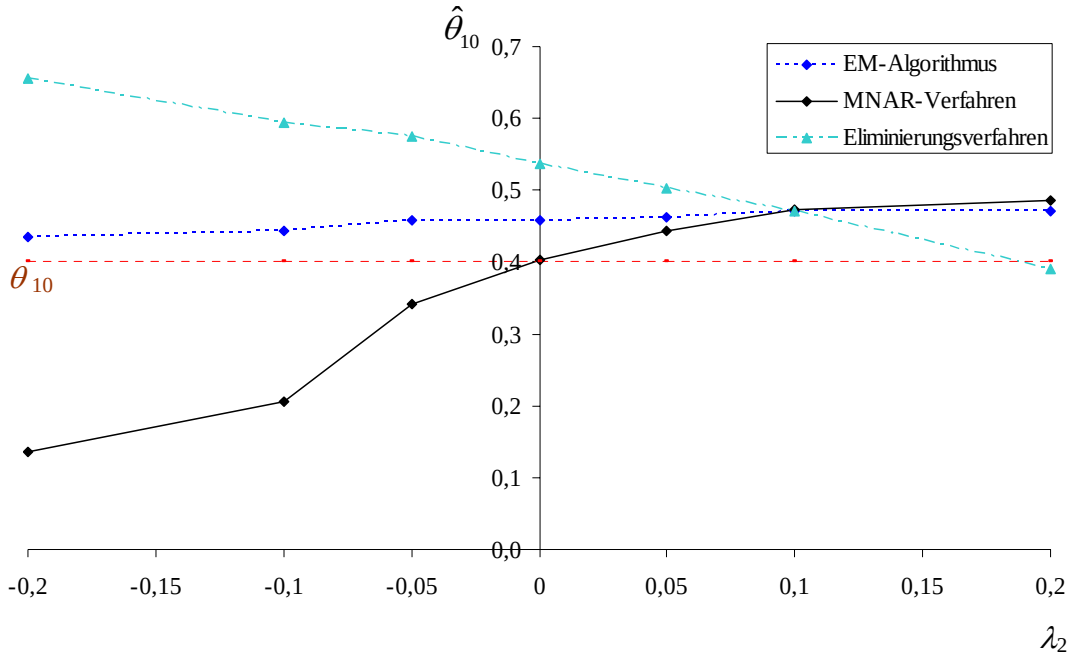


Abbildung 7.8: Schätzung des Parameters θ_{10} unter Variierung von λ_2

[214] Vgl. Bedingte Ausfallwahrscheinlichkeiten in (7.4).

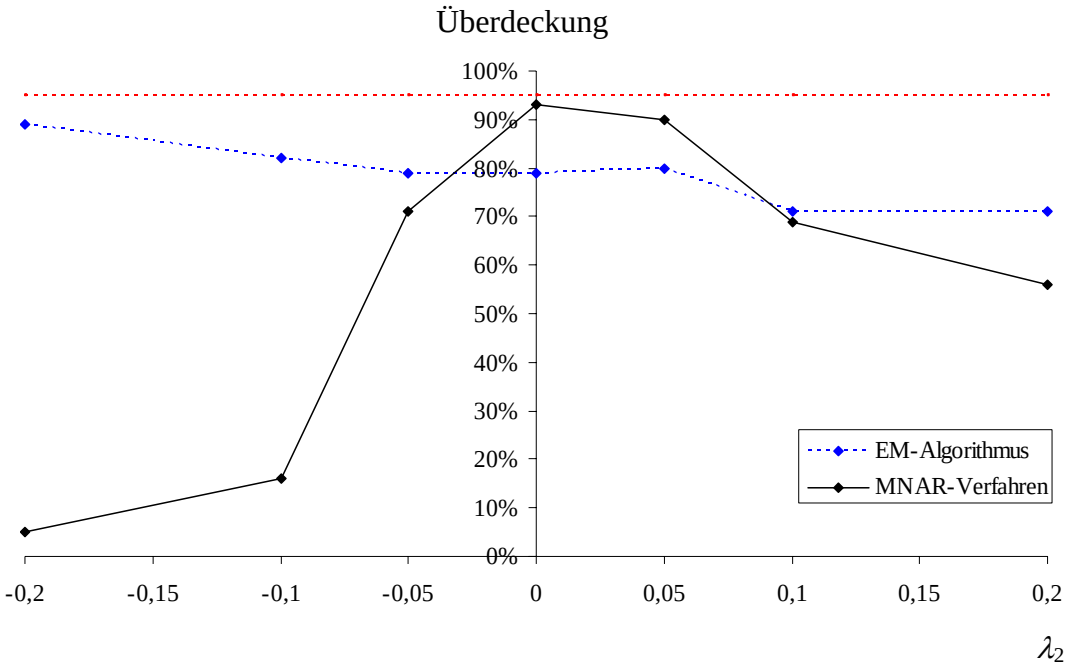


Abbildung 7.9: Überdeckung unter Variierung von λ_2

Die Auswertungen der Parameterschätzwerte und der Konfidenzintervalle zeigen, dass die MNAR-basierte Methode bereits bei einem geringen Zusammenhang zwischen R und Y_1 gegeben Y_2 $\left(|\lambda_2| \geq 0{,}05\right)$ zu verzerrten Ergebnissen führt.[215] Das MNAR-basierte Verfahren erweist sich somit als wenig robust gegenüber der Verletzung der bedingten Unabhängigkeitsannahme

$R_2 \perp\!\!\!\perp Y_1 \mid Y_2$.

Aus Abbildung 7.9 ist weiterhin ersichtlich, dass der EM-Algorithmus nicht von dem steigenden Informationsgehalt $\left(|\lambda_2| > 0\right)$ der Zufallsvariable Y_1 bezüglich des Datenausfalls profitiert. Dies ist durch den zugrunde liegenden Ausfallmechanismus zu erklären, der durch die beiden folgenden Graphen dargestellt werden kann.

[215] Vgl. Anhang A.8.

Abbildung 7.10: Graphische Darstellung des zugrunde liegenden Ausfallmechanismus ($\lambda_1 = -0{,}2, \lambda_2 \neq 0$)

Der Ausfallmechanismus ist vom Typ MNAR, wobei die Zufallsvariable Y_2 durch die Festlegung von λ_1 = -0,2 einen relativ großen Einfluss auf R_2 besitzt[216] und sich die Graphen in Abbildung 7.10 deutlich von den graphischen Modellierungen unter der MAR-Annahme[217] unterscheiden. Hieraus resultieren auch die gravierenden Abweichungen der weiteren Parameterschätzer, so dass die Anwendung des EM-Algorithmus unter diesem Ausfallmechanismus nicht vertretbar ist.

Die gesamte Betrachtung aller Parameterschätzer im Anhang A.10 zeigt, dass eine Behandlung durch das Eliminierungsverfahren ebenfalls zu verzerrten Ergebnissen führt. Es ist somit festzuhalten, dass die in der Praxis gebräuchlichen Methoden in Form des EM-Algorithmus und des Eliminierungsverfahrens nicht in der Lage sind, fehlende Werte unter den untersuchten, nicht ignorierbaren Ausfallmechanismen unverzerrt zu behandeln. Ist die Plausibilität der MAR-Annahme nicht gegeben, muss die Parameterschätzung mittels MNAR-basierter Methoden durchgeführt werden. Wie die Simulationsergebnisse zeigen, sind diese Verfahren jedoch stark an nicht überprüfbare Restriktionen bzw. (Unabhängigkeits-)Annahmen gebunden, so dass eine Sensitivitätsanalyse bezüglich der Schätzungen zwingend notwendig ist.

[216] Vgl. Abbildung 7.1.
[217] Vgl. Anhang A.9.

7.3 Untersuchung bei stetigen Variablen

7.3.1 Allgemeiner Simulationsaufbau

Im vorangegangenen Kapitel wurden die Auswirkungen auf die Parameterschätzungen dargestellt, falls die statistische Behandlung von fehlenden Werten einer diskreten Zufallsvariable durch Verfahren erfolgt, die zwingend die Erfüllung der MAR- bzw. MCAR-Annahme voraussetzen und der zugrunde liegende Ausfallmechanismus vom Typ MNAR ist. Um die Verzerrung der Schätzer im stetigen Fall beurteilen zu können, sei das in Kapitel 6.4.2.3 diskutierte bivariate Pattern-Mixture Modell der Ausgangspunkt der folgenden Simulationsstudie. In diesem Ansatz wurden zwei stetige Zufallsvariablen Y_1 und Y_2 betrachtet, wobei lediglich die letztgenannte Variable vom Datenausfall betroffen ist. Die Grundlage dieses Pattern-Mixture Modells bilden die folgenden Verteilungsannahmen:

$$\begin{aligned} &\left(Y_1, Y_2 \mid R_2 = r\right) \sim N\left(\mu^{(r)}, \Sigma^{(r)}\right) \qquad (r = 0, 1) \\ &R_2 \sim B(1, \varepsilon_1) \end{aligned} \tag{7.5}$$

Wie bereits in der vorangegangenen Untersuchung von diskreten Zufallsvariablen gelte die bedingte Unabhängigkeitsbeziehung

$$R_2 \perp\!\!\!\perp Y_1 \mid Y_2 \tag{7.6}$$

in dieser Simulationsstudie und der Ausfallmechanismus sei vom Typ MNAR. Das Ziel dieser Untersuchung besteht in der Bestimmung des Ausmaßes, in dem die Schätzer für den Erwartungswert μ_2 der Zufallsvariable Y_2 verzerrt sind, falls die Gültigkeit der MAR- bzw. MCAR-Annahme im Pattern-Mixture Modell irrtümlich vorausgesetzt wird. In Kapitel 6.4.2.3 wurde gezeigt, dass die bedingte Unabhängigkeitsbeziehung (7.6) erfüllt und der Ausfallmechanismus vom Typ MNAR ist, falls die Koeffizienten der Regressionen

$$Y_1^{(r)} = c^{(r)} + d^{(r)} Y_2^{(r)} + V^{(r)} \quad r = 0{,}1 \qquad \left(Y_j^{(r)} := Y_j \mid R = r \quad \forall j = 1{,}2\right)$$

und die Residualvarianzen übereinstimmen:[218]

[218] Vgl. Herleitung in (6.77)-(6.79).

$$d^{(0)} = d^{(1)} \text{ , } c^{(0)} = c^{(1)}, \sigma_{11\cdot 2}^{(0)} = \sigma_{11\cdot 2}^{(1)} \tag{7.7}$$

Dementsprechend werden *q* Realisationen der bivariat normalverteilten Zufallsvariablen $(Y_1^{(0)}, Y_2^{(0)})$ mit den Parametern

$$\mu_1^{(0)} = 0 \quad , \quad \mu_2^{(0)} = 0 \quad , \quad \sigma_{11}^{(0)} = 1 \quad , \quad \sigma_{22}^{(0)} = 1 \quad , \quad \sigma_{12}^{(0)} = 0{,}5 \tag{7.8}$$

im Pattern *r* = 0 gezogen, während die Simulation der (*n*-*q*) Werte der Zufallsvariablen $(Y_1^{(1)}, Y_2^{(1)})$ unter der Restriktion (7.7) erfolgt. Wegen

$$d^{(0)} = \frac{\sigma_{12}^{(0)}}{\sigma_{22}^{(0)}} = 0{,}5 \quad , \quad c^{(0)} = \mu_1^{(0)} - d^{(0)}\mu_2^{(0)} = 0 \quad , \quad \sigma_{11\cdot 2}^{(0)} = \sigma_{11}^{(0)} - \frac{\left(\sigma_{12}^{(0)}\right)^2}{\sigma_{22}^{(0)}} = 0{,}75$$

müssen die Parameter im Pattern *r* = 1 in der folgenden Weise voneinander abhängig sein, damit die Identität in (7.7) gewährleistet ist:

$$\text{(I) } \sigma_{12}^{(1)} = 0{,}5\sigma_{22}^{(1)} \qquad \text{(II) } \mu_2^{(1)} = 2\mu_1^{(1)} \qquad \text{(III) } \sigma_{11}^{(1)} = 0{,}75 + \frac{\left(\sigma_{12}^{(1)}\right)^2}{\sigma_{22}^{(1)}} \tag{7.9}$$

Um diese Bedingungen einzuhalten, gilt für die Kovarianzmatrizen in den beiden Pattern

$$\Sigma^{(0)} = \Sigma^{(1)}, \tag{7.10}$$

während der – aus den Daten schätzbare – Parameter $\mu_1^{(1)}$ in der Studie variiert wird.[219] Aufgrund des Simulationsaufbaus ist der interessierende Erwartungswert μ_2 von dem Parameter $\mu_1^{(1)}$ in der folgenden Weise abhängig:[220]

$$\mu_2 = (1-\varepsilon_1)\mu_2^{(0)} + \varepsilon_1\mu_2^{(1)} = \varepsilon_1\mu_2^{(1)} = 2\varepsilon_1\mu_1^{(1)} \tag{7.11}$$

Ausgehend von einer unvollständig beobachteten Stichprobe ist der Erwartungswert μ_2 mittels

$$\hat{\mu}_2 = (1-\hat{\varepsilon}_1)\hat{\mu}_2^{(0)} + \hat{\varepsilon}_1\hat{\mu}_2^{(1)} \tag{7.12}$$

zu schätzen. Während die Schätzer

---

[219] Aufgrund von Bedingung (II) in (7.9) wird gleichzeitig der Parameter $\mu_2^{(1)}$ in Abhängigkeit von $\mu_1^{(1)}$ festgelegt: $\mu_2^{(1)} = 2\mu_1^{(1)}$

[220] Vgl. (6.73), (7.8) und Bedingung (II) in (7.9).

$$\hat{\varepsilon}_1 = \frac{n-q}{n} \tag{7.13}$$

und

$$\hat{\mu}_2^{(0)} = \bar{y}_2^{(0)} = \frac{1}{q}\sum_{i=1}^{q} y_{i2} \tag{7.14}$$

unter allen Ausfalltypen (MCAR, MAR und MNAR) gleich sind, ergeben sich durch die Mechanismen unterschiedliche Schätzungen für $\mu_2^{(1)}$. Anhand der Behandlung der fehlenden Werte durch die entsprechenden Verfahren (Eliminierungsverfahren, Pattern-Mixture Modell unter MAR, Pattern-Mixture Modell unter MNAR) sowie der anschließenden Schätzung des Erwartungswerts μ_2 soll herausgefunden werden, inwieweit die Schätzwerte unter verschiedenen Simulationsbedingungen (in Form von unterschiedlichen Werten des Parameters $\mu_1^{(1)}$) verzerrt sind.

Um den Einfluss des Ausfalltyps allgemein zu verdeutlichen, sind in den folgenden Abbildungen die Dichtefunktionen von Y_2 dargestellt, die sich aus dem (tatsächlich zugrunde liegenden) MNAR-Ausfallmechanismus und der MAR-Annahme im Pattern-Mixture Modell ergeben. Dabei ist für jeden der beiden Ausfallmechanismen die Dichtefunktion von Y_2 unter den in (7.8) und (7.10) festgelegten Parameterwerten sowie $\mu_1^{(1)} = 2$, $\varepsilon_1 = 0{,}2$ angegeben.

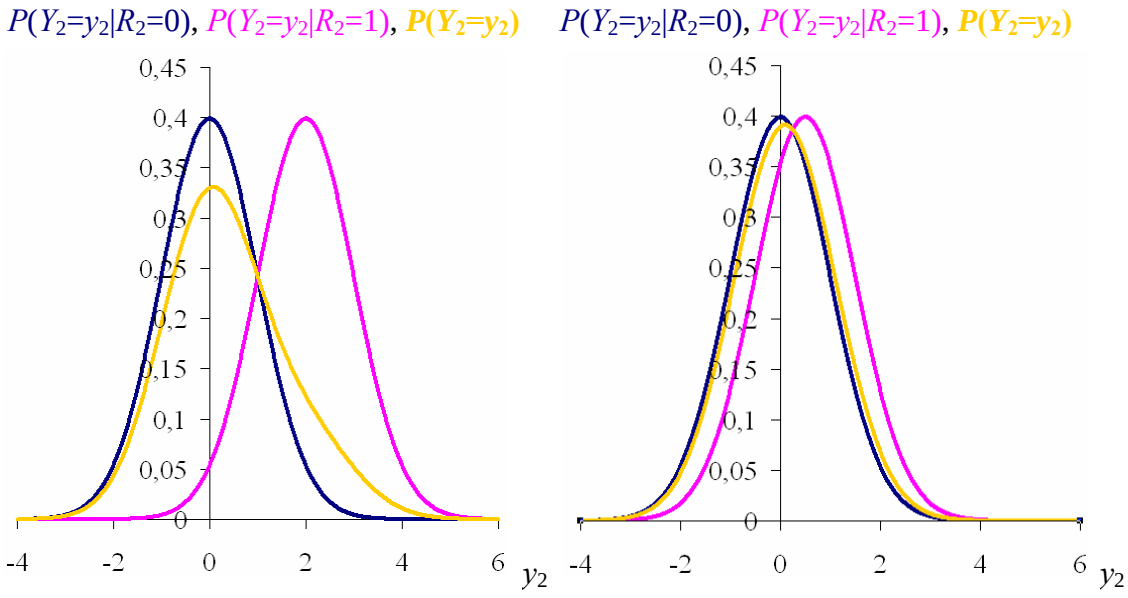


Abbildung 7.11: Dichtefunktionen von Y_2 unter einem Ausfallmechanismus vom Typ MNAR (linke Abb.) bzw. MAR (rechte Abb.)

Ein weiterer Faktor, der das Ausmaß der Verzerrung von $\hat{\mu}_2$ bestimmt, ist der Anteil der fehlenden Werte (*n*-*q*) am gesamten Stichprobenumfang *n*. Durch dessen Variierung ist zu überprüfen, ab welcher Ausfallquote die Schätzwerte der einzelnen Behandlungsmethoden von dem wahren Erwartungswert μ_2 signifikant abweichen. Weiterhin ist zu untersuchen, wie robust das Pattern-Mixture Modell für den Ausfalltyp MNAR unter anderen nicht ignorierbaren Ausfallmechanismen ist. Unter dieser Zielsetzung werden mehrere Datensimulationen unter den Restriktionen

$$d^{(1)} = \lambda\, d^{(0)}\,,\; c^{(1)} = c^{(0)}, \sigma_{11\cdot 2}^{(1)} = \sigma_{11\cdot 2}^{(0)} \quad (\lambda \neq 1)$$

und verschiedenen Werten des Parameters λ durchgeführt. Die Auswertung der Parameterschätzungen von dem Pattern-Mixture Modell soll Aufschluss darüber geben, inwieweit die Anwendungsmöglichkeiten des Verfahrens durch die zugrunde liegenden Annahmen bzw. Restriktionen begrenzt werden.

7.3.2 Durchführung und Ergebnisse der Simulation

Bereits im Kapitel 6.4.2.4 wurde anhand eines Beispiels festgestellt, dass sich die Schätzungen des Erwartungswerts μ_2 unter den beiden Ausfallmechanismen MAR bzw. MNAR erheblich voneinander unterscheiden können. Dieses Beispiel wird im Weiteren vertieft, indem unter verschiedenen Parameterwerten von $\mu_1^{(1)}$ und unter Berücksichtigung der bedingten Unabhängigkeitsbeziehung

$$R_2 \perp\!\!\!\perp Y_1 \mid Y_2 \tag{7.15}$$

die Simulation von 100 Stichproben vom Umfang $n = 100$ erfolgt. Die Simulation basiert auf den Parametervorgaben in (7.8) und (7.10), während der Wert des Parameters ε_1 durch

$$\varepsilon_1 = 0{,}2$$

festgelegt wurde. Die folgende Abbildung stellt für jede der untersuchten Behandlungsmethoden den durchschnittlichen Schätzwert $\hat{\mu}_2$ der 100 Stichproben dar. Zusätzlich ist der wahre Wert des Parameters μ_2 unter den verschiedenen Parameterwerten von $\mu_1^{(1)}$ angegeben, welcher sich aus dem Zusammenhang

$$\mu_2 = 2\varepsilon_1\mu_1^{(1)} = 0{,}4\mu_1^{(1)}$$

ergibt.[221]

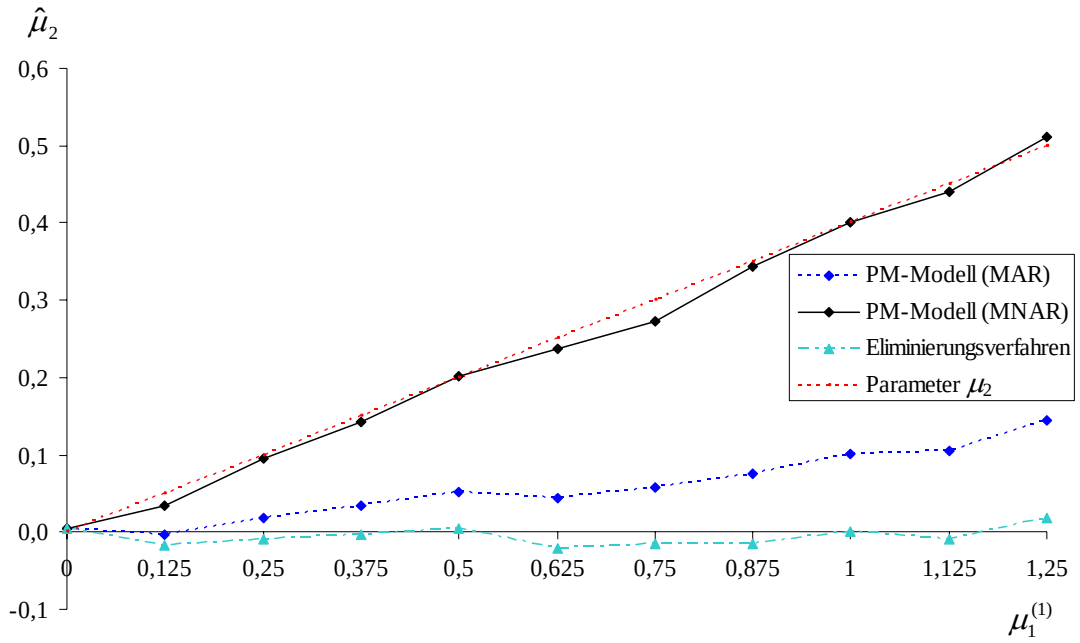


Abbildung 7.12: Schätzung von μ_2 bei Anwendung verschiedener Behandlungsmethoden und unterschiedlichen Parameterwerten von $\mu_1^{(1)}$

Aus den in Abbildung 7.12 dargestellten Simulationsergebnissen wird ersichtlich, dass die Abweichungen des Schätzers $\hat{\mu}_2$ im Pattern-Mixture Modell, welches auf einem MNAR-Ausfallmechanismus und der Gültigkeit der bedingten Unabhängigkeitsbeziehung (7.15) beruht,[222] vom wahren Parameter μ_2 lediglich von geringem Ausmaß sind. Wie bereits im vorangegangenen Kapitels erläutert entsprechen die Annahmen dieses Modells dem tatsächlich zugrunde liegenden Ausfallmechanismus, so dass die Abweichungen allein auf den Stichprobenfehler zurückzuführen sind.[223] Das Pattern-Mixture Modell unter der MAR-Annahme führt lediglich für den Parameterwert $\mu_1^{(1)} = 0$ zu einem unverzerrten Schätzer von μ_2, da in diesem Fall die

[221] Vgl. Formel (7.11).
[222] Dieses Modell wird in Abbildung 7.12 sowie den weiteren Darstellungen als PM-Modell (MNAR) bezeichnet.
[223] Vgl. Anhang A.11. Die Schätzer von μ_2 sind unter keinem der untersuchten Parameterwerte ($0 \leq \mu_1^{(1)} \leq 1{,}25$) verzerrt.

Parameter in den beiden Pattern übereinstimmen:

$$\mu_j^{(0)} = \mu_j^{(1)} = 0 \quad , \quad \Sigma^{(0)} = \Sigma^{(1)} = \begin{pmatrix} 1 & 0{,}5 \\ 0{,}5 & 1 \end{pmatrix} \quad \forall j = 1{,}2$$

Durch diese Identität der Parameter in den Pattern $r = 0$ und $r = 1$ ist die MCAR-Annahme (und somit auch die MAR-Annahme) erfüllt, da der Datenausfall weder von der Zufallsvariable Y_2 noch von Y_1 abhängig ist. Aufgrund dessen sind auch die Schätzungen von μ_2, die sich aus der Anwendung der drei Behandlungsmethoden ergeben, für $\mu_1^{(1)} = 0$ unverzerrt.[224]

Wird das Pattern-Mixture Modell unter der MAR-Annahme angewendet, so weisen die Schätzungen von μ_2 im Fall von $\mu_1^{(1)} > 0$ einen Bias auf, da dieses Modell den Zusammenhang[225]

$$\mu_2^{(1)} = a^{(0)} + b^{(0)} \mu_1^{(1)} = \left(\mu_1^{(0)} - \frac{\sigma_{12}^{(0)}}{\sigma_{11}^{(0)}} \mu_2^{(0)} \right) + \frac{\sigma_{12}^{(0)}}{\sigma_{11}^{(0)}} \mu_1^{(1)} = 0{,}5 \mu_1^{(1)}$$

und somit auch die Gültigkeit von

$$\mu_2 = (1 - \varepsilon_1) \mu_2^{(0)} + \varepsilon_1 \mu_2^{(1)} = 0{,}2 \mu_2^{(1)} = 0{,}1 \mu_1^{(1)} \tag{7.16}$$

voraussetzt. Der zugrunde liegende Ausfallmechanismus ist jedoch vom Typ MNAR, und im Gegensatz zu (7.16) gilt tatsächlich der Zusammenhang $\mu_2 = 0{,}4 \mu_1^{(1)}$.[226]

Für $\mu_1^{(1)} > 0$ wird somit die Schätzung des interessierenden Parameters μ_2 zu gering ausgewiesen, wenn in inkorrekter Weise von der Erfüllung der MAR-Annahme ausgegangen wird. Die systematische Unterschätzung von μ_2 durch das Pattern-Mixture Modell, welches die Gültigkeit der MAR-Annahme voraussetzt, kann somit auch innerhalb der Simulation festgestellt werden.[227]

Die unter den verschiedenen Simulationsbedingungen

$$0 < \mu_1^{(1)} \leq 1{,}25$$

---

[224] Vgl. Anhang A.11.
[225] Vgl. Formeln (6.65) und (6.67) des bivariaten Pattern-Mixture Modells unter der MAR-Annahme in Kapitel 6.4.2.2.
[226] Vgl. Formel (7.11).
[227] Vgl. Abbildung 7.12.

ermittelten Parameterschätzer $\hat{\mu}_2$ sind bei Anwendung des Eliminierungsverfahrens besonders stark verzerrt, weil diese Methode auf der MCAR-Annahme beruht und μ_2 mittels

$$\hat{\mu}_2 = \hat{\mu}_2^{(0)}$$

für jeden beliebigen Wert des Parameters $\mu_1^{(1)}$ geschätzt wird. Da die Simulation in der Weise durchgeführt wurde, dass der Zusammenhang $\mu_2 = 0{,}4\mu_1^{(1)}$ zwischen den Parametern μ_2 und $\mu_1^{(1)}$ unter allen Konstellationen besteht, sind bei steigenden Werten von $\mu_1^{(1)}$ höhere Abweichungen zwischen dem Schätzer $\hat{\mu}_2$ und dem wahren Wert des Parameters μ_2 zu beobachten, wenn die unvollständigen Datensätze eliminiert wurden. Das Eliminierungsverfahren erweist sich somit in dieser Simulationsstudie als eine ungeeignete Methode zur Behandlung von fehlenden Werten.

Die Verzerrung der Parameterschätzung, die aus der Pattern-Mixture Modellierung unter der MAR Annahme resultiert, kann ebenfalls anhand des Anteils der Konfidenzintervalle, die den wahren Parameter μ_2 enthalten, beurteilt werden.[228] Während im Fall der korrekten Modellierung durch den MNAR-Mechanismus kein Bias festgestellt werden konnte und die Überdeckung unter allen Simulationsbedingungen deutlich über 90% lag,[229] führt die Umsetzung des Pattern-Mixture Modells auf Basis der MAR-Annahme für alle untersuchten Parameterwerte $\mu_1^{(1)} > 0$ zu einer Überdeckung, die weniger als 90% beträgt. Wie bereits bei der Untersuchung der Parameterschätzwerte $\hat{\mu}_2$ (Abbildung 7.12) festgestellt wurde, wird die Verzerrung mit steigenden Parameterwerten $\mu_1^{(1)}$ größer. Als Ergebnis ist festzuhalten, dass eine auf der MAR-Annahme basierende Behandlung von fehlenden Werten einer stetigen Zufallsvariable zu einem erheblichen Bias von Schätzungen führen kann, wenn der zugrunde liegende Ausfallmechanismus vom Typ MNAR ist.

---

[228] Das Eliminierungsverfahren wurde aufgrund der geringeren effektiven Stichprobengröße q nicht berücksichtigt, da die Intervalle eine wesentlich höhere Breite aufweisen und die Aussagekraft der Überdeckung beeinträchtigt wird.

[229] Vgl. Überdeckung des PM-Modells (MNAR) in Abbildung 7.13.

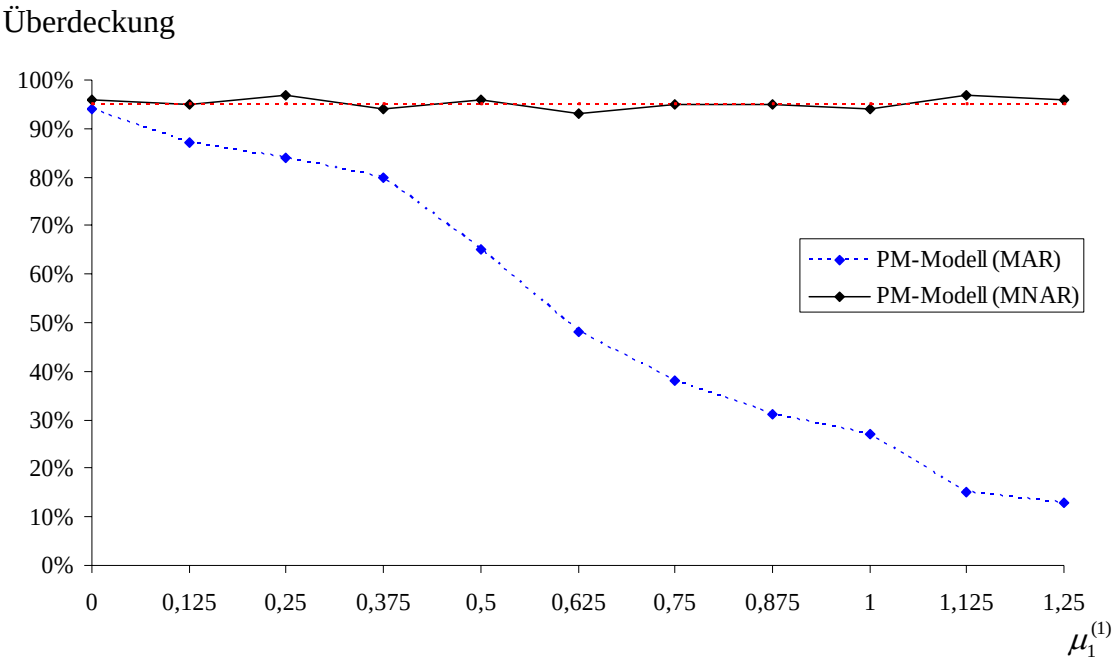


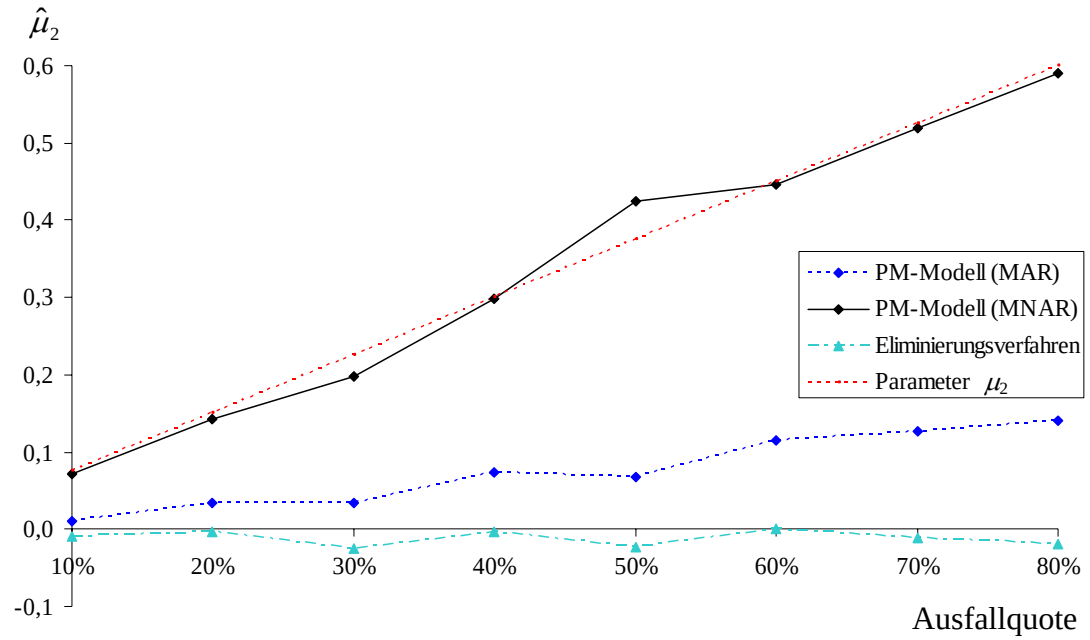
Abbildung 7.13: Überdeckung bei Anwendung des Pattern-Mixture Modells unter der MAR- bzw. MNAR-Annahme und unterschiedlichen Parameterwerten von $\mu_1^{(1)}$

Die Ausfallquote $\frac{n-q}{n} \cdot 100\%$ stellt eine weitere Größe dar, welche die Schätzung des Erwartungswerts der Zufallsvariable Y_2 beeinflusst.[230] Im Weiteren wird diese Quote unter der Simulationsbedingung $\mu_1^{(1)} = 0{,}375$ verändert, um das aus den unterschiedlichen Behandlungsmethoden resultierende Ausmaß der Verzerrung von $\hat{\mu}_2$ zu ermitteln. Die folgende Abbildung zeigt die wesentlichen Ergebnisse dieser Simulation in Form der Parameterschätzungen sowie der Überdeckungen unter verschiedenen Ausfallquoten. Weiterhin ist der wahre Wert des Parameters μ_2 angegeben, der sich mittels

$$\mu_2 = (1-\varepsilon_1)\mu_2^{(0)} + \varepsilon_1\mu_2^{(1)} = \varepsilon_1\mu_2^{(1)} = 2\varepsilon_1\mu_1^{(1)} = 0{,}75\varepsilon_1$$

berechnen lässt.

[230] Vgl. Formel (7.12). Der Anteil fehlender Werte (*n*-*q*) am Stichprobenumfang *n* ist dabei der Schätzer des Parameters ε_1 (vgl. (7.13)).



Überdeckung

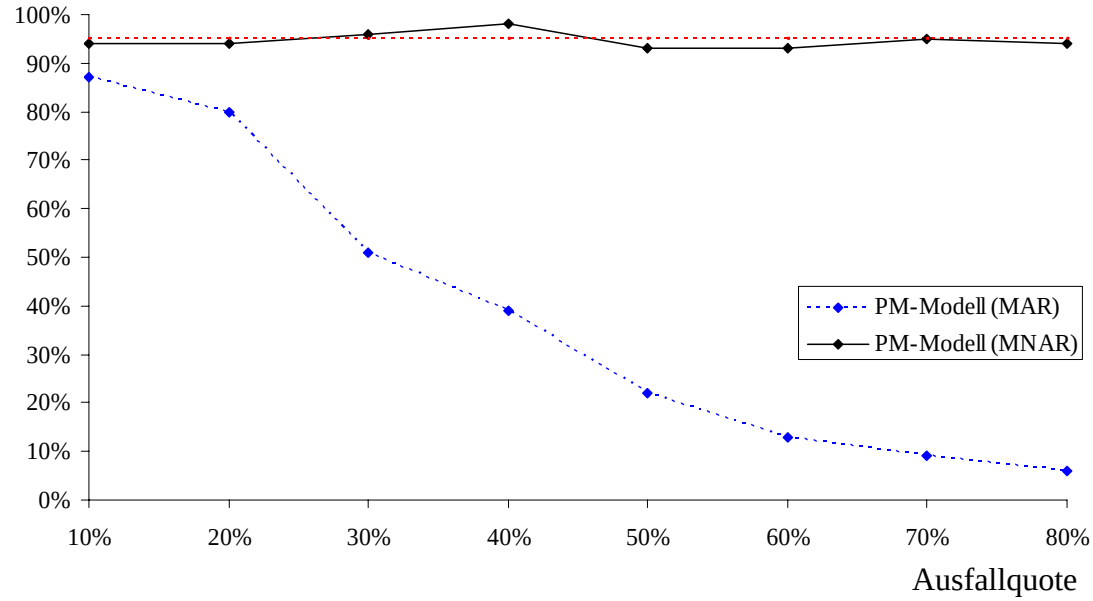


Abbildung 7.14: Parameterschätzung und Überdeckung unter verschiedenen Ausfallquoten

Das Pattern-Mixture Modell unter MNAR, welches auf der der Gültigkeit der bedingten Unabhängigkeitsbeziehung

$R_2 \perp\!\!\!\perp Y_1 \mid Y_2$

basiert, führt unter allen untersuchten Ausfallquoten zu unverzerrten Schätzungen des Parameters μ_2. Dies wird sowohl durch die Untersuchung des Bias anhand des Standardfehlers des Schätzers[231] als auch durch die Überdeckung bestätigt, die in allen Fällen über 90% liegt. Hingegen erweisen sich die MCAR- und MAR-basierten Verfahren unter dem zugrunde liegenden Ausfallmechanismus selbst bei einem geringen Anteil fehlender Werte als nicht geeignet, den Datenausfall zu behandeln und eine unverzerrte Schätzung des interessierenden Parameters zu ermöglichen. Die Ergebnisse des Eliminierungsverfahrens weichen von den wahren Werten des Parameters μ_2 gravierend ab, da die Annahme

$$\mu_2^{(0)} = \mu_2^{(1)} = \mu_2,$$

auf der das MCAR-basierte Verfahren beruht, nicht erfüllt ist und sich dies insbesondere bei einer hohen Ausfallquote negativ auf die Schätzungen auswirkt.

Die bisherigen Simulationen zeigten, dass die Behandlung von fehlenden Werten durch MCAR- und MAR-basierten Methoden zu erheblich verzerrten Schätzern führt, wenn der Ausfallmechanismus vom Typ MNAR ist und die bedingte Unabhängigkeit von R_2 und Y_1 gegeben Y_2 gilt. In Kapitel 6.4.2.3 wurde nachgewiesen, dass diese bedingte Unabhängigkeit die Gültigkeit der Restriktionen

$$d^{(0)} = d^{(1)}\ ,\ c^{(0)} = c^{(1)}, \sigma_{11\cdot 2}^{(0)} = \sigma_{11\cdot 2}^{(1)}$$

impliziert.[232] In einer weiteren Datensimulation wurde untersucht, wie robust das bivariate Pattern-Mixture Modell unter MNAR ist, indem von einer dieser Restriktionen mittels

$$d^{(1)} = \lambda\, d^{(0)} \quad (\lambda \neq 1)$$

abgewichen wurde. Die folgenden Abbildungen zeigen die Parameterschätzungen und Überdeckungen der Behandlungsmethoden für verschiedene Werte des Parameters λ, die sich bei einer Ausfallquote von 20% und dem Parameterwert $\mu_2^{(1)} = 0{,}75$ ergeben.

---

[231] Vgl. Anhang A.12.
[232] Vgl. Herleitung in (6.77)-(6.79).

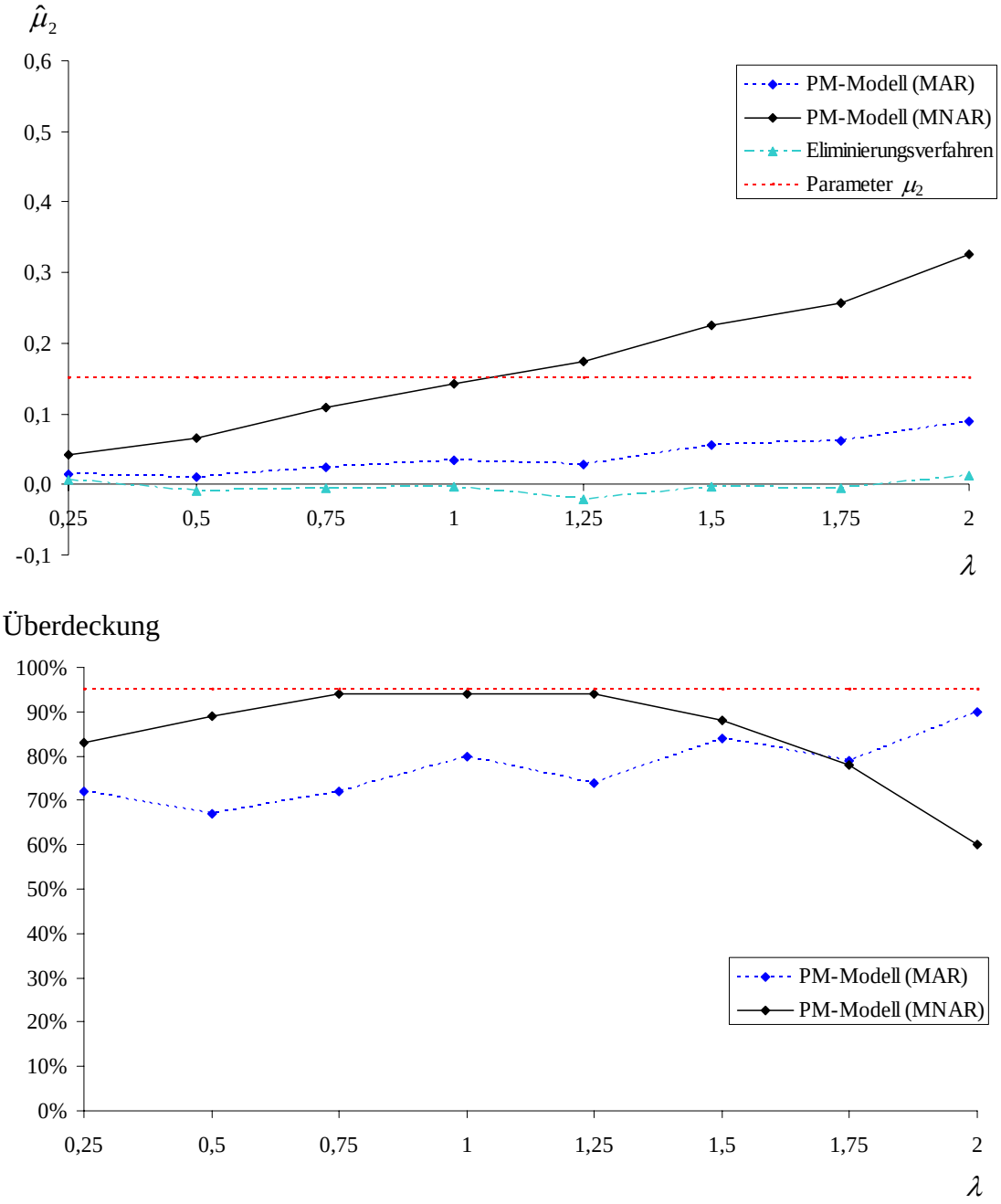


Abbildung 7.15: Parameterschätzungen und Überdeckung unter Variierung von λ

Die Parameterschätzungen des Pattern-Mixture Modells unter MNAR, welches auf der Gültigkeit der Restriktionen

$$d^{(0)} = d^{(1)}\,,\; c^{(0)} = c^{(1)},\; \sigma_{11\cdot 2}^{(0)} = \sigma_{11\cdot 2}^{(1)}$$

beruht, sind im Fall von $0{,}75 \leq \lambda \leq 1{,}25$ unverzerrt.[233] Hingegen weisen die Schätzungen unter den weiteren untersuchten Werten von λ einen Bias auf, da sich in diesen Fällen die Abweichungen von den Restriktionen des Modells zu deutlich auf die Schätzungen auswirken. Unter dem Aspekt, dass lediglich eine Restriktion verändert wurde, ist somit festzustellen, dass eine zweckmäßige Behandlung von fehlenden Werten durch das Pattern-Mixture Modell relativ stark an die Einhaltung von dessen Restriktionen gebunden ist. Da der Ausfallmechanismus in der Simulation vom Typ MNAR ist, sind die Ergebnisse, die aus der MCAR- und MAR-Annahme resultieren, unter allen Bedingungen verzerrt. Eine Anwendung des Eliminierungsverfahrens bzw. des Pattern-Mixture Modells unter der MAR-Annahme ist somit auch in dieser Simulationsstudie nicht gerechtfertigt.

7.4 Fazit

Die in diesem Kapitel durchgeführten Simulationen zeigen, dass die Problematik des nicht ignorierbaren Datenausfalls durch die Anwendung von MCAR- bzw. MAR-basierten Methoden nicht gelöst werden kann. Sowohl bei der Betrachtung von diskreten als auch von stetigen Zufallsvariablen sind die Parameterschätzungen unter nahezu allen untersuchten Szenarien verzerrt, wenn eine Behandlung von fehlenden Werten durch diese Verfahren erfolgt. Die Simulationsergebnisse lassen somit die von Schafer/Graham (2002) vertretene These, dass sich in vielen Fällen eine Verletzung der MAR-Annahme nur unwesentlich auf die Schätzungen auswirkt,[234] zumindest zweifelhaft erscheinen. Für das Ausmaß der Verzerrung unter dieser Annahme ist ausschlaggebend, wie stark der Ausfall von der unvollständig beobachteten Variable abhängig und wie hoch die Ausfallquote ist. Kann dieser Zusammenhang nicht aus plausiblen Gründen (z.B. durch Auswertung einer nachträglichen Befragung von nicht antwortenden Personen) ausgeschlossen werden, ist es insbesondere bei einer hohen Ausfallquote zwingend notwendig, den nicht ignorierbaren Ausfallmechanismus zu modellieren. Die Simulationsergebnisse verdeutlichen, dass eine inkorrekte Spezifizierung des nicht ignorierbaren Ausfallmechanismus ebenfalls zu verzerrten Schätzern führt. Unter diesem Gesichtspunkt kann die Auffassung zahlreicher Auto-

[233] Vgl. Anhang A.13.
[234] Vgl. Schafer/Graham (2002), S. 152.

ren, die eine Sensitivitätsanalyse der Parameterschätzungen für unerlässlich halten,[235] bestätigt werden.

[235] Vgl. Foster/Fang (2004); Allison (2002), S. 78; Toutenburg et al. (2004), S. 34; Scharfstein et al. (1999); Copas/Eguchi (2001).

8 Zusammenfassung und Ausblick

Die vorliegende Arbeit befasste sich mit der Problematik von fehlenden Werten, deren Ausfall nicht durch die beobachteten Daten erklärt werden kann. Im Mittelpunkt der Ausführungen standen Verfahren, die eine Behandlung von fehlenden Daten unter dieser Konstellation ermöglichen. Aufgrund des nicht ignorierbaren Ausfallmechanismus beruhen die entsprechenden Verfahren zwangsläufig auf Annahmen, die anhand der beobachteten Daten nicht überprüfbar sind. Dies rechtfertigt jedoch nicht die in der Praxis anzutreffende Präferenz von MAR-basierten Methoden, da diese Verfahren ebenfalls die Erfüllung einer kritischen Annahme (MAR) voraussetzen. Vielmehr sind Ansätze von verstärkter Bedeutung, die der bestehenden Unsicherheit bezüglich des Ausfallmechanismus Rechnung tragen und eine Beurteilung der Sensitivität von Schätzungen erlauben. In dieser Hinsicht besteht ein erheblicher Forschungsbedarf und ein wesentliches Ziel der Arbeit war es, diese Lücke zu füllen. Vor diesem Hintergrund wurde in den Kapiteln 6.2.2, 6.4.2.4 und 6.4.3 der theoretische Rahmen für problemspezifische Sensitivitätsanalysen geschaffen, der die Grundlage für die anschließenden praktischen Umsetzungen bildete. Die Ergebnisse der Analysen lieferten bereits Anhaltspunkte über die Auswirkungen der MAR-Annahme auf die Güte der Schätzungen, falls der Ausfallmechanismus nicht ignorierbar ist. Insofern konnte an dieser Stelle die Notwendigkeit einer Sensitivitätsbetrachtung im Kontext der Behandlung von fehlenden Werten unterstrichen werden.

Aus theoretischer Sicht kann eine Sensitivitätsanalyse im Rahmen der Selection und Pattern-Mixture Modellierung, die in den Kapiteln 6.3 und 6.4 diskutiert wurden, erfolgen. Die Selection Modelle beruhen dabei auf Verteilungsannahmen, die sowohl für das Datenmodell als auch für den nicht ignorierbaren Ausfallmechanismus zu treffen sind. In der Arbeit wurde dargestellt, wie dieser Ansatz für eine univariat normalverteilte Zufallsvariable und verschiedene Modellierungen des Ausfallmechanismus (Schwellenwert-Modellierung, Logit-Modellierung) praktisch realisiert werden kann. Dabei erwies sich die Maximierung der entsprechenden Likelihood-Funktion in den betrachteten Fällen als problematisch, so dass die Schätzer in den Selection Modellen mittels iterativer Näherungsverfahren bestimmt wurden. Unter diesem Aspekt wird deutlich, dass eine an die Selection Modellierung geknüpfte Sensitivitätsanalyse nur unter erheblichem Aufwand umgesetzt werden kann. Ferner

ist die Güte der Schätzungen stark von den Verteilungsannahmen des Datenmodells abhängig, so dass die Selection Modelle nur eingeschränkt die Möglichkeit eröffnen, fehlende Werte unter nicht ignorierbaren Ausfallmechanismen geeignet zu behandeln.

In der Arbeit wurde mit den Pattern-Mixture Modellen ein vielversprechender Ansatz ausführlich diskutiert, der sowohl aus theoretischer Sicht als auch unter dem Gesichtspunkt der praktischen Realisierbarkeit mit einer erforderlichen Sensitivitätsbetrachtung vereinbar ist. Der methodische Vorteil des Ansatzes liegt dabei in einer Verdeutlichung derjenigen modellspezifischen Parameter, die anhand der beobachteten Daten nicht identifizierbar sind. Durch eine Variierung der Restriktionen, die für diese Parameter zu formulieren sind, kann in unkomplizierter Weise eine Beurteilung der Sensitivität von Schätzungen erfolgen. Darüber hinaus wurde eine Erweiterung dieses Ansatzes in Form der Pattern-Set Mixture Modellierung vorgenommen, die sich insbesondere der Problematik von Unit Nonresponse unter nicht ignorierbaren Ausfallmechanismen widmet. In dem Zusammenhang ist gezeigt worden, wie verschiedene Mechanismen für Item und Unit Nonresponse in dem erweiterten Modell Berücksichtigung finden können.

Ein weiterer Schwerpunkt der Arbeit befasste sich mit der Frage, inwieweit Schätzungen verzerrt sind, falls MAR-basierte Methoden zur Anwendung kommen und der Ausfallmechanismus nicht ignorierbar ist. Die Simulationsstudien in Kapitel 7 lieferten die Erkenntnis, dass bereits bei einem geringen Einfluss der unvollständig beobachteten Variable auf den Datenausfall sowie einer niedrigen Ausfallquote die Schätzungen einen Bias aufweisen. Insofern kann auch die in der Literatur vertretene Auffassung, dass eine Verletzung der MAR-Annahme die Güte von Schätzungen in vielen Fällen nur unwesentlich beeinflusst,[236] nicht geteilt werden. Weiterhin wurde deutlich, dass sich die Behandlungsverfahren für MNAR als nicht robust gegenüber der Verletzung ihrer (Unabhängigkeits-)Annahmen bzw. Restriktionen erweisen. Die bereits in den theoretischen Kapiteln manifestierte Erkenntnis, dass eine Durchführung von Sensitivitätsanalysen im Kontext der Behandlung von fehlenden Werten erforderlich ist, findet somit ihre Bestätigung in den Simulationsstudien.

[236] Vgl. Schafer/Graham (2002), S. 152.

In dieser Arbeit beschränkte sich die Betrachtung auf die Methodik der Behandlungsverfahren sowie auf Datensimulationen, die Aufschluss über deren Eignung bei Vorliegen von nicht ignorierbaren Ausfallmechanismen geben konnten. Im Rahmen weiterführender Studien ist zu überprüfen, ob die wesentlichen Ergebnisse der Arbeit auch aus empirischer Sicht bestätigt werden können. Generell besteht in dieser Hinsicht ein erheblicher Forschungsbedarf, da eine Validierung der (MAR- und MNAR-basierten) Verfahren auf empirischer Ebene bisher vernachlässigt wurde.[237] Aus Anwendersicht ist die Entwicklung von Statistik-Software, die Sensitivitätsanalysen in Verbindung mit der Pattern-Mixture Modellierung allgemein zugänglich macht, voranzutreiben. Dies würde auf der einen Seite zu einer verbreiteten Akzeptanz der Modelle beitragen und andererseits dem Anwender die Problematik, welche den nicht ignorierbaren Ausfallmechanismen immanent ist, eröffnen.

[237] Vgl. Schnell (1997), S. 245ff.

Anhang

A.1

Der Erwartungswert von Q unter Vollständigkeit ist durch

$$E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) = \int Q\, P(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) dQ$$

definiert. Bei wiederholtem Ziehen von Werten für $\mathbf{Y}_{mis}$ aus der Verteilung $P(\mathbf{Y}_{mis} \mid \mathbf{y}_{obs})$ erhält man

$$\begin{aligned} E\left[E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right] &= \int \left[\int Q\, P(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) dQ\right] P(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}) d\mathbf{y}_{mis} \\ &= \int\int Q\, P(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) P(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}) dQ\, d\mathbf{y}_{mis} \end{aligned} \tag{A.1}$$

als Erwartungswert von Q.

Aufgrund des Satzes von Bayes gilt

$$P(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) = \frac{P(\mathbf{y}_{obs}, \mathbf{y}_{mis} \mid Q) P(Q)}{P(\mathbf{y}_{obs}, \mathbf{y}_{mis})}$$

und somit

$$P(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) P(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}) = \frac{P(\mathbf{y}_{obs}, \mathbf{y}_{mis} \mid Q) P(Q)}{P(\mathbf{y}_{obs})}. \tag{A.2}$$

Durch Einsetzen von (A.2) in (A.1) erhält man den bedingten Erwartungswert von Q gegeben die beobachteten Daten:

$$\begin{aligned} E\left[E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right] &= \int\int Q \frac{P(Q)}{P(\mathbf{y}_{obs})} P(\mathbf{y}_{mis}, \mathbf{y}_{obs} \mid Q) dQ\, d\mathbf{y}_{mis} \\ &= \int Q \frac{P(Q)}{P(\mathbf{y}_{obs})} \left[\int P(\mathbf{y}_{mis}, \mathbf{y}_{obs} \mid Q)\, d\mathbf{y}_{mis}\right] dQ \\ &= \int Q \frac{P(Q)}{P(\mathbf{y}_{obs})} P(\mathbf{y}_{obs} \mid Q) dQ \\ &= \int Q \frac{P(Q) P(\mathbf{y}_{obs}, Q)}{P(\mathbf{y}_{obs}) P(Q)} dQ \\ &= \int Q P(Q \mid \mathbf{y}_{obs})\, dQ \\ &= E(Q \mid \mathbf{y}_{obs}) \end{aligned}$$

A.2

Für die bedingte Varianz von Q gegeben $\mathbf{y}$

$$\mathrm{Var}(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) = E(Q^2 \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) - (E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}))^2$$

ist der bedingte Erwartungswert gegeben die Beobachtungen $\mathbf{y}_{obs}$ zu bilden:

$$\begin{aligned} \mathrm{E}\left[\mathrm{Var}(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right] &= E\left[E(Q^2 \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right] - E\left[(E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}))^2 \mid \mathbf{y}_{obs}\right] \\ &= E(Q^2 \mid \mathbf{y}_{obs}) - E\left[(E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}))^2 \mid \mathbf{y}_{obs}\right] \end{aligned} \tag{A.3}$$

Aufgrund von

$$\begin{aligned} \mathrm{Var}(E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}) &= E\left[(E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}))^2 \mid \mathbf{y}_{obs}\right] - \left(E\left[E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right]\right)^2 \\ &= E\left[(E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}))^2 \mid \mathbf{y}_{obs}\right] - \left(E(Q \mid \mathbf{y}_{obs})\right)^2 \end{aligned}$$

erhält man durch Einsetzen in (A.3)

$$\begin{aligned} \mathrm{E}\left[\mathrm{Var}(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right] &= E(Q^2 \mid \mathbf{y}_{obs}) - \mathrm{Var}(E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}) - \left(E(Q \mid \mathbf{y}_{obs})\right)^2 \\ &= \mathrm{Var}(Q \mid \mathbf{y}_{obs}) - \mathrm{Var}(E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}), \end{aligned}$$

und die Varianz von Q gegeben die beobachteten Daten ist

$$\mathrm{Var}(Q \mid \mathbf{y}_{obs}) = E\left[\mathrm{Var}(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right] + \mathrm{Var}\left[E(Q \mid \mathbf{y}_{obs}, \mathbf{y}_{mis}) \mid \mathbf{y}_{obs}\right].$$

A.3

Ist der Ausfallmechanismus ignorierbar und ist somit die MAR-Annahme erfüllt, gilt nach Formel (4.3) für die Dichte der beobachteten Daten

$$P_{\theta,\psi}(\mathbf{r},\mathbf{y}_{obs}) = P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})\, P_{\theta}(\mathbf{y}_{obs}) .$$

Durch Einsetzen in die a posteriori Verteilung

$$P(\theta,\psi \mid \mathbf{y}_{obs},\mathbf{r}) = \frac{P_{\theta,\psi}(\mathbf{r},\mathbf{y}_{obs})P(\theta,\psi)}{\int\int P_{\theta,\psi}(\mathbf{r},\mathbf{y}_{obs})P(\theta,\psi)d\theta\, d\psi}$$

erhält man

$$P(\theta,\psi \mid \mathbf{y}_{obs},\mathbf{r}) = \frac{P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P_{\theta}(\mathbf{y}_{obs})P(\theta,\psi)}{\int\int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P_{\theta}(\mathbf{y}_{obs})P(\theta,\psi)d\theta\, d\psi},$$

und wegen Voraussetzung (II) der (bayesianischen) Definition 4.2 von Ignorierbarkeit (Unabhängigkeit der a priori Verteilungen von θ und ψ) gilt

$$P(\theta,\psi \mid \mathbf{y}_{obs},\mathbf{r}) = \frac{P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P_{\theta}(\mathbf{y}_{obs})P(\theta)P(\psi)}{\int\int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P_{\theta}(\mathbf{y}_{obs})P(\theta)P(\psi)d\theta\, d\psi}$$

$$= \frac{P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P_{\theta}(\mathbf{y}_{obs})P(\theta)P(\psi)}{\int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P(\psi)d\psi \int P_{\theta}(\mathbf{y}_{obs})P(\theta)\, d\theta}.$$

Die marginale a posteriori Verteilung von θ gegeben $\mathbf{y}_{obs}$ und $\mathbf{r}$, $P(\theta \mid \mathbf{y}_{obs},\mathbf{r})$, entspricht dann der marginalen a posteriori Verteilung von θ gegeben $\mathbf{y}_{obs}$, $P(\theta \mid \mathbf{y}_{obs})$:

$$P(\theta \mid \mathbf{y}_{obs},\mathbf{r}) = \int P(\theta,\psi \mid \mathbf{y}_{obs},\mathbf{r})d\psi$$

$$= \frac{\int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P_{\theta}(\mathbf{y}_{obs})P(\theta)P(\psi)d\psi}{\int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P(\psi)d\psi \int P_{\theta}(\mathbf{y}_{obs})P(\theta)\, d\theta}$$

$$= \frac{P_{\theta}(\mathbf{y}_{obs})P(\theta)\int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P(\psi)d\psi}{\int P_{\psi}(\mathbf{r} \mid \mathbf{y}_{obs})P(\psi)d\psi \int P_{\theta}(\mathbf{y}_{obs})P(\theta)\, d\theta}$$

$$= \frac{P_{\theta}(\mathbf{y}_{obs})P(\theta)}{\int P_{\theta}(\mathbf{y}_{obs})P(\theta)\, d\theta}$$

$$= P(\theta \mid \mathbf{y}_{obs})$$

A.4 Schätzung des Parameters θ_{10} unter Variierung von λ_1

	Parameter λ_1								
	-0,4	-0,3	-0,2	-0,1	-0,05	0	0,05	0,1	0,2	0,266
EM-Algorithmus										
Überdeckung	43,0%	63,0%	78,9%	92,0%	94,0%	96,0%	90,0%	87,0%	54,0%	38,0%
∅ Intervalllänge	0,1959	0,1958	0,1954	0,1945	0,1931	0,1925	0,1901	0,1891	0,1822	0,1780
∅ Schätzer	0,5052	0,4816	0,4601	0,4368	0,4134	0,4038	0,3791	0,3686	0,3190	0,2931
Varianz des Schätzers	0,0027	0,0028	0,0025	0,0023	0,0023	0,0020	0,0025	0,0022	0,0029	0,0025
Bias	JA	JA	JA	JA	NEIN	NEIN	NEIN	JA	JA	JA
MNAR-Verfahren										
Überdeckung	96,0%	93,0%	95,0%	94,0%	95,0%	96,0%	95,0%	97,0%	96,0%	94,0%
∅ Intervalllänge	0,1926	0,1919	0,1922	0,1926	0,1918	0,1925	0,1917	0,1925	0,1916	0,1924
∅ Schätzer	0,4058	0,3995	0,4021	0,4070	0,3975	0,4038	0,3961	0,4042	0,3956	0,4047
Varianz des Schätzers	0,0022	0,0026	0,0024	0,0024	0,0023	0,0020	0,0025	0,0020	0,0026	0,0024
Bias	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN
Eliminierungsverfahren										
∅ Schätzer	0,6759	0,6001	0,5340	0,4727	0,4309	0,4038	0,3633	0,3375	0,2650	0,2276
Varianz des Schätzers	0,0032	0,0038	0,0031	0,0026	0,0025	0,0020	0,0023	0,0018	0,0018	0,0013
Bias	JA	JA	JA	JA	JA	NEIN	JA	JA	JA	JA

A.5 p-Werte des χ^2-Anpassungstests unter Variierung von λ_1

($\theta_{00} = 0{,}2$, $\theta_{10} = 0{,}4$, $\theta_{01} = 0{,}3$, $\theta_{11} = 0{,}1$)

	λ_1	$\hat{\theta}_{00}$	$\hat{\theta}_{10}$	$\hat{\theta}_{01}$	$\hat{\theta}_{11}$	p-Wert
EM-Algorithmus	-0,4	0,4948	0,5052	0,0000	0,0000	0,00000
	-0,3	0,3944	0,4816	0,1030	0,0210	0,00000
	-0,2	0,3173	0,4601	0,1778	0,0448	0,00124
	-0,1	0,2536	0,4368	0,2415	0,0681	0,26917
	-0,05	0,2216	0,4134	0,2792	0,0858	0,89023
	0	0,1945	0,4038	0,3011	0,1006	0,99928
	0,05	0,1785	0,3791	0,3253	0,1171	0,83899
	0,1	0,1516	0,3686	0,3443	0,1355	0,34310
	0,2	0,1181	0,3190	0,3872	0,1757	0,00412
	0,266	0,0928	0,2931	0,4032	0,2109	0,00002

	λ_1	$\hat{\theta}_{00}$	$\hat{\theta}_{10}$	$\hat{\theta}_{01}$	$\hat{\theta}_{11}$	p-Wert
MNAR-Verfahren	-0,4	0,1947	0,4058	0,3001	0,0994	0,99909
	-0,3	0,1985	0,3995	0,2989	0,1031	0,99969
	-0,2	0,2010	0,4021	0,2941	0,1028	0,99919
	-0,1	0,2008	0,4070	0,2943	0,0979	0,99878
	-0,05	0,1962	0,3975	0,3046	0,1017	0,99932
	0	0,1945	0,4038	0,3011	0,1006	0,99928
	0,05	0,2022	0,3961	0,3016	0,1001	0,99984
	0,1	0,1962	0,4042	0,2997	0,0999	0,99967
	0,2	0,2042	0,3956	0,3011	0,0991	0,99952
	0,266	0,2008	0,4047	0,2952	0,0993	0,99956

	λ_1	$\hat{\theta}_{00}$	$\hat{\theta}_{10}$	$\hat{\theta}_{01}$	$\hat{\theta}_{11}$	p-Wert
Eliminierungsverfahren	-0,4	0,3241	0,6759	0,0000	0,0000	0,00000
	-0,3	0,2980	0,6001	0,0758	0,0261	0,00000
	-0,2	0,2667	0,5340	0,1476	0,0517	0,00078
	-0,1	0,2331	0,4727	0,2208	0,0734	0,19797
	-0,05	0,2127	0,4309	0,2672	0,0892	0,85107
	0	0,1945	0,4038	0,3011	0,1006	0,99928
	0,05	0,1855	0,3633	0,3388	0,1125	0,77756
	0,1	0,1640	0,3375	0,3739	0,1246	0,25576
	0,2	0,1366	0,2650	0,4502	0,1481	0,00094
	0,266	0,1129	0,2276	0,4936	0,1659	0,00000

A.6 Schätzung des Parameters θ_{10} unter verschiedenen Ausfallquoten

	Ausfallquote					
	20%	30%	40%	50%	60%	70%
EM-Algorithmus						
Überdeckung	87,0%	86,0%	82,0%	76,0%	61,0%	47,0%
∅ Intervalllänge	0,1947	0,1947	0,1952	0,1953	0,1959	0,1961
∅ Schätzer	0,4430	0,4437	0,4552	0,4582	0,4856	0,5017
Varianz des Schätzers	0,0025	0,0025	0,0024	0,0026	0,0026	0,0023
Bias	JA	JA	JA	JA	JA	JA
MNAR-Verfahren						
Überdeckung	93,0%	94,0%	94,0%	97,0%	95,0%	96,0%
∅ Intervalllänge	0,1918	0,1915	0,1920	0,1913	0,1923	0,1921
∅ Schätzer	0,3986	0,3956	0,4005	0,3907	0,4055	0,3991
Varianz des Schätzers	0,0026	0,0027	0,0025	0,0022	0,0028	0,0020
Bias	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN
Eliminierungsverfahren						
∅ Schätzer	0,5003	0,5105	0,5319	0,5488	0,6078	0,6658
Varianz des Schätzers	0,0030	0,0031	0,0031	0,0028	0,0036	0,0028
Bias	JA	JA	JA	JA	JA	JA

A.7 p-Werte des χ^2-Anpassungstests unter verschiedenen Ausfallquoten *AQ* ($\theta_{00} = 0{,}2$, $\theta_{10} = 0{,}4$, $\theta_{01} = 0{,}3$, $\theta_{11} = 0{,}1$)

	AQ	$\hat{\theta}_{00}$	$\hat{\theta}_{10}$	$\hat{\theta}_{01}$	$\hat{\theta}_{11}$	p-Wert
EM-Algorithmus	20%	0,2759	0,4430	0,2243	0,0568	0,06805
	30%	0,2961	0,4437	0,2108	0,0494	0,01614
	40%	0,3209	0,4552	0,1823	0,0416	0,00109
	50%	0,3541	0,4582	0,1530	0,0347	0,00002
	60%	0,3900	0,4856	0,1039	0,0205	0,00000
	70%	0,4983	0,5017	0,0000	0,0000	0,00000

	AQ	$\hat{\theta}_{00}$	$\hat{\theta}_{10}$	$\hat{\theta}_{01}$	$\hat{\theta}_{11}$	p-Wert
MNAR-Verfahren	20%	0,1921	0,3986	0,3081	0,1012	0,99663
	30%	0,1965	0,3956	0,3104	0,0975	0,99678
	40%	0,2019	0,4005	0,3013	0,0963	0,99946
	50%	0,2046	0,3907	0,3025	0,1022	0,99797
	60%	0,1937	0,4055	0,3002	0,1006	0,99878
	70%	0,2001	0,3991	0,2982	0,1026	0,99981

	AQ	$\hat{\theta}_{00}$	$\hat{\theta}_{10}$	$\hat{\theta}_{01}$	$\hat{\theta}_{11}$	p-Wert
Eliminierungsver-fahren	20%	0,2412	0,5003	0,1945	0,0639	0,03884
	30%	0,2538	0,5105	0,1792	0,0565	0,01042
	40%	0,2686	0,5319	0,1512	0,0483	0,00079
	50%	0,2875	0,5488	0,1225	0,0412	0,00003
	60%	0,2909	0,6078	0,0759	0,0255	0,00000
	70%	0,3342	0,6658	0,0000	0,0000	0,00000

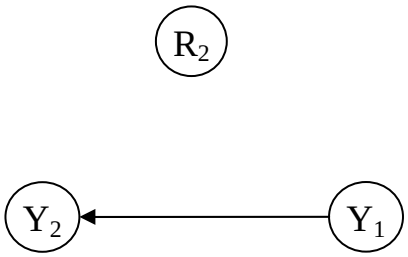
A.8 Schätzung des Parameters θ_{10} unter Variierung von λ_2

	Parameter λ_2						
	-0,2	-0,1	-0,05	0	0,05	0,1	0,2
EM-Algorithmus							
Überdeckung	89,0%	82,0%	79,0%	79,0%	80,0%	71,0%	71,0%
∅ Intervalllänge	0,1944	0,1948	0,1954	0,1953	0,1957	0,1957	0,1956
∅ Schätzer	0,4343	0,4443	0,4586	0,4592	0,4618	0,4707	0,4703
Varianz des Schätzers	0,0021	0,0024	0,0023	0,0025	0,0017	0,0024	0,0026
Bias	JA	JA	JA	JA	JA	JA	JA
MNAR-Verfahren							
Überdeckung	5,0%	16,0%	71,0%	93,0%	90,0%	69,0%	56,0%
∅ Intervalllänge	0,1046	0,1442	0,1852	0,1922	0,1951	0,1958	0,1959
∅ Schätzer	0,1362	0,2051	0,3421	0,4029	0,4439	0,4730	0,4867
Varianz des Schätzers	0,1059	0,0121	0,0036	0,0026	0,0017	0,0023	0,0027
Bias	JA	JA	JA	NEIN	JA	JA	JA
Eliminierungsverfahren							
∅ Schätzer	0,6548	0,5946	0,5742	0,5356	0,5024	0,4718	0,3894
Varianz des Schätzers	0,0021	0,0029	0,0027	0,0032	0,0021	0,0030	0,0029
Bias	JA	JA	JA	JA	JA	JA	NEIN

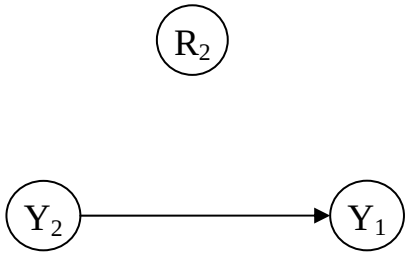
A.9 Graphische Darstellung der Ausfallmechanismen unter verschiedenen Annahmen

MCAR-Annahme:

(I)

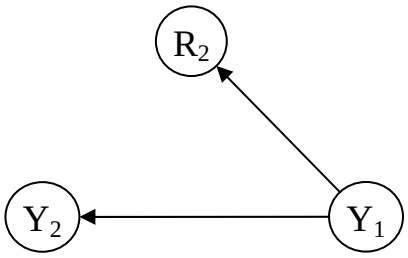


(II)

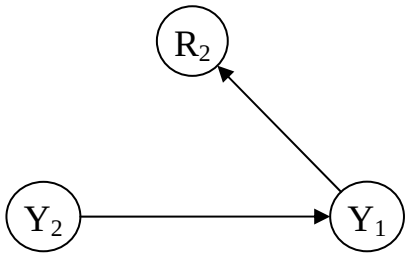


MAR-Annahme:

(I)

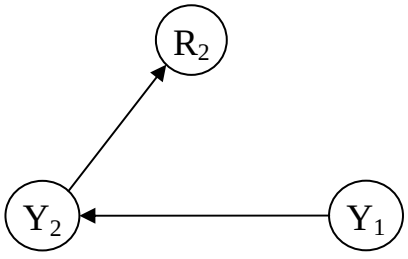


(II)

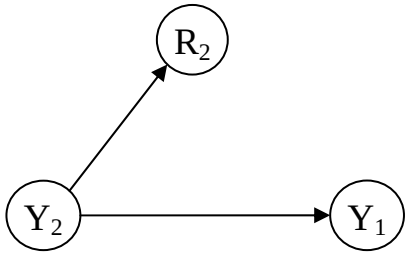


Ausfalltyp MNAR und bedingte Unabhängigkeitsannahme $R_2 \perp\!\!\!\perp Y_1 \mid Y_2$:

(I)



(II)



A.10 p-Werte des χ2-Anpassungstests unter Variierung von λ_2

(θ_{00} = 0,2, θ_{10} = 0,4, θ_{01} = 0,3, θ_{11} = 0,1)

	λ_2	$\hat{\theta}_{00}$	$\hat{\theta}_{10}$	$\hat{\theta}_{01}$	$\hat{\theta}_{11}$	p-Wert
EM-Algorithmus	-0,2	0,4004	0,4343	0,1059	0,0594	0,00000
	-0,1	0,3473	0,4443	0,1572	0,0512	0,00013
	-0,05	0,3254	0,4586	0,1707	0,0453	0,00062
	0	0,3157	0,4592	0,1819	0,0432	0,00147
	0,05	0,3105	0,4618	0,1919	0,0358	0,00175
	0,1	0,2983	0,4707	0,1985	0,0325	0,00280
	0,2	0,2957	0,4703	0,2139	0,0201	0,00214

	λ_2	$\hat{\theta}_{00}$	$\hat{\theta}_{10}$	$\hat{\theta}_{01}$	$\hat{\theta}_{11}$	p-Wert
MNAR-Verfahren	-0,2	0,0111	0,0362	0,4952	0,4575	0,00000
	-0,1	0,0748	0,2051	0,4297	0,2904	0,00000
	-0,05	0,1436	0,3421	0,3525	0,1618	0,06673
	0	0,1987	0,4029	0,2989	0,0995	0,99994
	0,05	0,2532	0,4439	0,2492	0,0537	0,17923
	0,1	0,3023	0,4730	0,1945	0,0302	0,00171
	0,2	0,4385	0,4867	0,0711	0,0037	0,00000

	λ_2	$\hat{\theta}_{00}$	$\hat{\theta}_{10}$	$\hat{\theta}_{01}$	$\hat{\theta}_{11}$	p-Wert
Eliminierungsver-fahren	-0,2	0,2033	0,6548	0,0525	0,0894	0,00000
	-0,1	0,2334	0,5946	0,1038	0,0682	0,00003
	-0,05	0,2437	0,5742	0,1257	0,0565	0,00013
	0	0,2641	0,5356	0,1502	0,0501	0,00085
	0,05	0,2849	0,5024	0,1742	0,0386	0,00160
	0,1	0,2987	0,4718	0,1972	0,0322	0,00255
	0,2	0,3454	0,3894	0,2487	0,0165	0,00035

A.11 Schätzung des Erwartungswerts μ_2 unter Variierung von $\mu_1^{(1)}$

	$\mu_1^{(1)}$										
	0	0,125	0,25	0,375	0,5	0,625	0,75	0,875	1	1,125	1,25
PM-Modell (MAR)											
Überdeckung	94,0%	87,0%	84,0%	80,0%	65,0%	48,0%	38,0%	31,0%	27,0%	15,0%	13,0%
∅ Intervalllänge	0,3746	0,3778	0,3803	0,3846	0,3899	0,4075	0,4173	0,4382	0,4485	0,4698	0,4849
∅ Schätzer	0,0040	-0,0032	0,0177	0,0337	0,0527	0,0446	0,0568	0,0759	0,1009	0,1054	0,1449
Varianz des Schätzers	0,0102	0,0116	0,0126	0,0108	0,0131	0,0112	0,0097	0,0127	0,0130	0,0122	0,0130
Bias	NEIN	JA	JA	JA	JA	JA	JA	JA	JA	JA	JA
PM-Modell (MNAR)											
Überdeckung	96,0%	95,0%	97,0%	94,0%	96,0%	93,0%	95,0%	95,0%	94,0%	97,0%	96,0%
∅ Intervalllänge	0,3920	0,3940	0,3998	0,4093	0,4222	0,4383	0,4571	0,4785	0,5020	0,5274	0,5544
∅ Schätzer	0,0038	0,0349	0,0961	0,1431	0,2008	0,2376	0,2734	0,3440	0,4004	0,4403	0,5113
Varianz des Schätzers	0,0089	0,0103	0,0098	0,0096	0,0106	0,0083	0,0080	0,0101	0,0101	0,0105	0,0104
Bias	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN
Eliminierungsverfahren											
∅ Schätzer	0,0037	-0,0166	-0,0101	-0,0037	0,0049	-0,0211	-0,0156	-0,0162	0,0001	-0,0086	0,0179
Varianz des Schätzers	0,0110	0,0124	0,0136	0,0118	0,0130	0,0117	0,0106	0,0129	0,0121	0,0125	0,0124
Bias	NEIN	JA	JA	JA	JA	JA	JA	JA	JA	JA	JA

A.12 Schätzung des Erwartungswerts μ_2 unter verschiedenen Ausfallquoten

	Ausfallquote							
	10%	20%	30%	40%	50%	60%	70%	80%
PM-Modell (MAR)								
Überdeckung	87,0%	80,0%	51,0%	39,0%	22,0%	13,0%	9,0%	6,0%
∅ Intervalllänge	0,3948	0,3846	0,3893	0,3795	0,3675	0,3658	0,3554	0,3453
∅ Schätzer	0,0107	0,0337	0,0332	0,0730	0,0670	0,1155	0,1271	0,1414
Varianz des Schätzers	0,0122	0,0108	0,0120	0,0142	0,0199	0,0233	0,0299	0,0350
Bias	JA	JA	JA	JA	JA	JA	JA	JA
PM-Modell (MNAR)								
Überdeckung	94,0%	94,0%	96,0%	98,0%	93,0%	93,0%	95,0%	94,0%
∅ Intervalllänge	0,4018	0,4093	0,4145	0,4176	0,4187	0,4176	0,4145	0,4093
∅ Schätzer	0,0708	0,1431	0,1973	0,2987	0,4237	0,4464	0,5194	0,5897
Varianz des Schätzers	0,0113	0,0096	0,0097	0,0080	0,0107	0,0126	0,0101	0,0121
Bias	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN	NEIN
Eliminierungsverfahren								
∅ Schätzer	-0,0093	-0,0037	-0,0225	-0,0034	-0,0222	0,0000	-0,0111	-0,0194
Varianz des Schätzers	0,0130	0,0118	0,0140	0,0146	0,0219	0,0229	0,0350	0,0381
Bias	JA	JA	JA	JA	JA	JA	JA	JA

A.13 Schätzung des Erwartungswerts μ_2 unter Variierung von λ

	λ							
	0,25	0,5	0,75	1	1,25	1,5	1,75	2
PM-Modell (MAR)								
Überdeckung	72,0%	67,0%	72,0%	80,0%	74,0%	84,0%	79,0%	90,0%
∅ Intervalllänge	0,3989	0,3972	0,3937	0,3846	0,3871	0,3837	0,3815	0,3822
∅ Schätzer	0,0149	0,0096	0,0243	0,0337	0,0285	0,0551	0,0616	0,0894
Varianz des Schätzers	0,0119	0,0126	0,0114	0,0108	0,0158	0,0119	0,0135	0,0101
Bias	JA	JA	JA	JA	JA	JA	JA	JA
PM-Modell (MNAR)								
Überdeckung	83,0%	89,0%	94,0%	94,0%	94,0%	88,0%	78,0%	60,0%
∅ Intervalllänge	0,4093	0,4093	0,4093	0,4093	0,4093	0,4093	0,4093	0,4093
∅ Schätzer	0,0424	0,0663	0,1095	0,1431	0,1745	0,2248	0,2566	0,3255
Varianz des Schätzers	0,0085	0,0094	0,0088	0,0096	0,0142	0,0112	0,0181	0,0136
Bias	JA	JA	NEIN	NEIN	NEIN	JA	JA	JA
Eliminierungsverfahren								
∅ Schätzer	0,0064	-0,0099	-0,0056	-0,0037	-0,0209	-0,0032	-0,0052	0,0131
Varianz des Schätzers	0,0129	0,0139	0,0125	0,0118	0,0167	0,0129	0,0127	0,0092
Bias	JA	JA	JA	JA	JA	JA	JA	JA

Literaturverzeichnis

Allison, P. D. (2002)

Missing Data; Sage Publications, Inc., Thousand Oaks, London, New Delhi.

Anderson, T. W. (1957)

Maximum Likelihood Estimates for the Multivariate Normal Distribution when some Observations are Missing; in: Journal of the American Statistical Association, Vol. 52, S. 200-203.

Baker, S. G.; Laird, N. M. (1988)

Regression Analysis for Categorical Variables with Outcome Subject to Nonignorable Nonresponse; in: *Journal of the American Statistical Association*, Vol. 83, S. 62-69.

Bamberg, G. (1972)

Statistische Entscheidungstheorie; Physica Verlag, Würzburg, Wien.

Bankhofer, U.; Praxmarer, S. (1998)

Zur Behandlung fehlender Daten in der Marktforschungspraxis; in: *Marketing – Zeitschrift für Forschung und Praxis*, Bd. 20, Heft 2, S. 109-118.

Cochran, W. G. (1977)

Sampling Techniques, Third Edition, John Wiley & Sons, New York.

Copas, J.; Eguchi, S. (2001)

Local Sensitivity Approximations for Selectivity Bias; in: *Journal of the Royal Statistical Society Series B*, Vol. 63, S. 871-895.

Crawford, S. L.; Tennstedt, S. L.; McKinlay, J. B. (1995)

A Comparison of Analytic Methods for Non-Random Missingness of Outcome Data; in: *Journal of Clinical Epidemiology*, Vol. 48, S. 209-219.

Daniels, M. J.; Hogan, J. W. (2000)

Reparameterizing the Pattern Mixture Model for Sensitivity Analysis under Informative Dropout; in: *Biometrics*, Vol. 56, S. 1241-1248.

David, M.; Little, R. J. A.; Samuhel, M. E.; Triest, R. K. (1986)

Alternative Methods for CPS Income Imputation; in: *Journal of the American Statistical Association*, Vol. 81, S. 29-41.

De Heer, W. (1999)

International Response Trends: Results of an International Survey; in: *Journal of Official Statistics*, Vol. 15, S. 129-142.

Dempster, A. P.; Laird, N. M.; Rubin, D. B. (1977)

Maximum Likelihood from Incomplete Data via the EM Algorithm; in: *Journal of the Royal Statistical Society Series B*, Vol. 39, S. 1-38.

Ekholm, A.; Skinner, C. (1998)

The Muscatine Children's Obesity Data Reanalysed Using Pattern Mixture Models; in: *Applied Statistics*, Vol. 47, S. 251-263.

Esser, H.; Grohmann, H.; Müller, W.; Schäfer, K. (1989)

Mikrozensus im Wandel; Metzler-Poeschel, Stuttgart.

Forster, J. J.; Smith, P. W. F. (1998)

Model-Based Inference for Categorical Survey Data Subject to Non-Ignorable Non-Response; in: *Journal of the Royal Statistical Society Series B*, Vol. 60, S. 57-70.

Foster, E. M.; Fang, G. Y. (2004)

Alternative Methods for Handling Attrition: An Illustration Using Data from the Fast Track Evaluation; in: *Evaluation Review*, Vol. 28, S. 434-464.

Glynn, R. J.; Laird, N. M.; Rubin, D. B. (1986)

Selection Modeling Versus Mixture Modeling with Nonignorable Nonresponse; in: Wainer, H. (ed.): *Drawing Inferences from Self-Selected Samples*, Springer Verlag, New York, S. 115-142.

Glynn, R. J.; Laird, N. M.; Rubin, D. B. (1993)

Multiple Impuation in Mixture Models for Nonignorable Nonresponse With Follow-ups; in: *Journal of the American Statistical Association*, Vol. 88, S. 984-993.

Graham, J. W.; Donaldson, S. I. (1993)

Evaluating Interventions With Differential Attrition: The Importance of Nonresponse Mechanisms and Use of Follow-Up Data; in: *Journal of Applied Psychology*, Vol. 78, S. 119-128.

Greene, W. H. (2000)

Econometric Analysis; Fourth Edition, Prentice Hall International, New York.

Greenlees, J. S.; Reece, W. S.; Zieschang, K. D. (1982)

Imputation of Missing Values When the Probability of Response Depends On the Variable Being Imputed; in: *Journal of the American Association*, Vol. 77, S. 251-261.

Groves, R.M. (1989)

Survey Errors and Survey Costs; John Wiley & Sons, New York.

Hartung, J. (1999)

Statistik; Oldenbourg Verlag, München, Wien.

Holland, P. W. (1986)

A Comment on Remarks by Rubin and Hartigan; in: Wainer, H. (ed.): *Drawing Inferences from Self-Selected Samples*, Springer Verlag, New York; S. 149-151.

Hübler, O. (1986)

Zufällig und systematisch fehlende Werte in linearen Regressionsmodellen; in: *Allgemeines Statistisches Archiv,* Bd. 70, S. 138-157.

Huisman, M. (1998)

Missing Data in Behavioral Science Research: Investigation of a Collection of Data Sets; in: *Kwantitatieve Methoden*, Vol. 57, S. 69-93.

Huisman, M. (1999)

Imputation of Missing Item Responses: Some Simple Techniques; in: Huisman, M. (Hrsg.): *Item Nonresponse: Occurrence, Causes, and Imputation of Missing Answers to Test Items*, DSWO Press, Leiden, S. 91-119.

Huisman, M.; Krol, B.; van Sonderen, E. (1999)

Handling Missing Data by Re-Approaching Nonrespondents; in: Huisman, M. (Hrsg.): *Item Nonresponse: Occurrence, Causes, and Imputation of Missing Answers to Test Items*, DSWO Press, Leiden, S. 47-62.

Jones, S. M.; Chromy, J. R. (1982)

Improved Variance Estimators Using Weighting Class Adjustments for Sample Survey Nonresponse; Proceedings of the Survey Research Methods Section, American Statistical Association.

Kastner, C. (2001)

Fehlende Werte bei korrelierten Beobachtungen, Dissertation; in: Toutenburg, H. (Hrsg.): *Anwendungsorientierte Statistik*, Peter Lang GmbH, Europäischer Verlag der Wissenschaften, Frankfurt am Main.

Kenward, M. G.; Molenberghs G.; Thijs, H. (2003)

Pattern-Mixture Models with Proper Time Dependence; in: *Biometrika*, Vol. 90, S. 53-71.

Landerman, L. R.; Land K. C.; Pieper, C. F. (1997)

An Empirical Evaluation of the Predictive Mean Matching Method for Imputing Missing Values; in: *Sociological Methods & Research*, Vol. 26, S. 3-33.

Lauritzen, S. L. (1996)

Graphical Models, Clarendon Press, Oxford.

Li, K.H. (1988)

Imputation using Markov Chains; in: *Journal of Statistical Computation and Simulation*, Vol. 30, S. 57-79.

Lillard, L.; Smith, J. P.; Welch, F. (1986)

What Do We Really Know about Wages? The Importance of Nonreporting and Census Imputation; in: *Journal of Political Economy*, Vol. 94, S. 489-506.

Little, R. J. A. (1988)

Missing-Data Adjustments in Large Surveys; in: *Journal of Business & Economic Statistics*, Vol. 6, S. 287-296.

Little, R. J. A. (1993)

Pattern-Mixture Models for Multivariate Incomplete Data; in: *Journal of the American Statistical Association*, Vol. 88, S. 125-134.

Little, R. J. A. (1994)

A Class of Pattern-Mixture Models for Normal Incomplete Data; in: *Biometrika*, Vol. 81, S. 471-483.

Little, R. J. A.; Rubin, D. B. (2002)

Statistical Analysis with Missing Data; Second Edition, John Wiley & Sons, New York.

Little, R. J. A.; Vartivarian, S. (2003)

On Weighting the Rates in Nonresponse Weights; in: *Statistics in Medicine*, Vol. 22, S. 1589-1599.

Molenberghs, G.; Michiels, B.; Kenward, M. G.; Diggle, P. J. (1998)

Monotone Missing Data and Pattern-Mixture Models; in: *Statistica Neerlandica*, Vol. 52, S. 153-161.

Nordheim, E. V. (1984)

Inference From Nonrandomly Missing Categorical Data: An Example From a Genetic Study of Turner´s Syndrom; in: *Journal of the American Statistical Association*, Vol. 79, S. 772-780.

Rässler, S. (2000)

Ergänzung fehlender Daten in Umfragen; in: *Jahrbücher für Nationalökonomie und Statistik*, Lucius & Lucius, Stuttgart, Bd. 220, Heft 1, S. 64-94.

Ronning, G. (1991)

Mikroökonometrie; Springer Verlag, Berlin [u.a.].

Rubin, D. B. (1976)

Inference and Missing Data; in: *Biometrica*, Vol. 63, S. 581-592.

Rubin, D. B. (1987)

Multiple Imputation for Nonresponse in Surveys; John Wiley & Sons, Inc., New York.

Rubin, D. B. (1996)

Multiple Imputation after 18+ Years; in: *Journal of the American Statistical Association*, Vol. 91, S. 473-489.

Rubin, D. B.; Schenker, N. (1986)

Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse; in: *Journal of the American Statistical Association*, Vol. 81, S. 366-374.

Runte, M. (1999)

Missing Values – Konzepte und statistische Literatur; Working Paper, Kiel.

Schafer, J. L. (1997)

Analysis of Incomplete Multivariate Data; Chapman & Hall, London.

Schafer, J. L. (1999)

Multiple Imputation: A Primer; in: *Statistical Methods in Medical Research*, Vol. 8, S. 3-15.

Schafer, J. L.; Graham, J. W. (2002)

Missing Data: Our View of the State of the Art; in: *Psychological Methods*, Vol. 7, S. 147-177.

Schafer, J. L.; Olsen, M. K. (1998)

Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst´s Perspective; Working Paper, Pennsylvania State University.

Scharfstein, D. O.; Rotnitzky, A.; Robins, J. M. (1999)

Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models; in: *Journal of the American Statistical Association*, Vol. 94, S. 1096-1120.

Schnell, R. (1986)

Missing-Data-Probleme in der empirischen Sozialforschung; Dissertation, Ruhr-Universität Bochum.

Schnell, R. (1997)

Nonresponse in Bevölkerungsumfragen; Leske + Budrich, Opladen.

Schnell, R. (2002)

Antworten auf Nonresponse, Manuskript zum Vortrag auf dem XXXVII. Kongress der deutschen Marktforschung am 7. Mai 2002 in Wolfsburg; Universität Konstanz.

Smith, P. W. F.; Skinner, C. J.; Clarke, P. S. (1999)

Allowing for Non-Ignorable Non-Response in the Analysis of Voting Intention Data; in: *Journal of the Royal Statistical Society Series C*, Vol. 48, S. 563-577.

Statistical Services (2000)

Handling Missing or Incomplete Data; Statistical Support, Devision of Research Consulting at ITS, University of Texas at Austin, http://www.utexas.edu/its/rc/answers/general/gen25.html.

Stolzenberg, R. M.; Relles, D. A. (1990)

Theory Testing in a World of Constrained Research Design; in: *Sociological Methods and Research*, Vol. 18, S. 395-415.

Storck, S.; Kastner, C.; Toutenburg, H. (2000)

Longitudinal Data with Dropouts: A Comparison of Pattern Mixture Models with Complete Case Analysis; Discussion Paper, Department für Statistik, Ludwig-Maximilians-Universität München.

Toutenburg H.; Heumann, C.; Nittner, T. (2004)

Statistische Methoden bei unvollständigen Daten; Diskussionspapier, Department für Statistik, Ludwig-Maximilians-Universität München.

Voß, W. (2000)

Taschenbuch der Statistik; Carl Hanser Verlag, München, Wien.

Wilhelm, Manfred G. (1996)

Analyse korrelierter Daten: Behandlung fehlender Werte und Modelldiagnostik; Dissertation, Institut für Mathematik der Medizinischen Universität zu Lübeck.

Lebenslauf

Name:	Thomas Lehmann
Geburtsdatum:	08.02.1974
Geburtsort:	Gotha

Bildungsweg

1980-1990	Polytechnische Oberschule GutsMuths, Waltershausen
1990-1992	Gymnasium „Salzmannschule“, Schnepfenthal
09/1993-09/1998	Studium der Betriebswirtschaftslehre an der Friedrich-Schiller-Universität Jena
⇨ 09/1998	Diplom (Note 1,99). Thema der Diplomarbeit: „Entwicklung und Untersuchung von Imputations-Algorithmen für Missing Values zur Datenvorbereitung für das Neuro-Fuzzy-System NEFCLASS“

Tätigkeiten

12/1992-12/1993, 08/1994-10/1994	Zivildienst
07/1996-10/1996	Praktikum in der Steuerberatungskanzlei Jannert, Gotha
11/1998-12/2000	Wissenschaftlicher Mitarbeiter in dem Projekt „Enhancing Competitiveness of Small and Medium Sized Enterprises via Innovation (ECOVIN)“
01/2001-09/2004	Wissenschaftlicher Mitarbeiter am Lehrstuhl für Wirtschafts- und Sozialstatistik, Universität Jena. Lehrstuhlinhaber: Prof. Dr. Kischka

Jena, 19.10.2004 _____

Erklärung gemäß § 4 Abs. 1, Pkt. 3 PromO

Hiermit erkläre ich,

1. dass mir die geltende Promotionsordnung bekannt ist;
2. dass die Dissertation von mir selbst angefertigt wurde und dass alle von mir benutzten Hilfsmittel und Quellen in der Arbeit angegeben wurden;
3. dass mich keine Personen bei der Auswertung des Materials sowie bei der Herstellung des Manuskriptes unterstützt haben;
4. dass die Hilfe eines Promotionsberaters nicht in Anspruch genommen wurde, und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen;
5. dass ich die Dissertation zuvor noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe;
6. dass ich weder die gleiche, noch eine in wesentlichen Teilen ähnliche, noch eine andere Abhandlung bei einer anderen Hochschule bzw. anderen Fakultät als Dissertation eingereicht habe.

Jena, 19.10.2004 _____