

Varianzschätzung unter Imputation und bei komplexen Stichprobendesigns

Dissertation

Universität Trier

Lehrstuhl für Wirtschafts- und
Sozialstatistik

Zur Erlangung des Grades:

Dr. rer. pol

Eingereicht bei:

Herrn Professor Dr. Ralf Münnich

Herrn PD Dr. Siegfried Gabler

im Oktober 2015

vorgelegt von:

Christian Bruch

geboren am 12.06.1983

in Saarbrücken

Danksagung

Zunächst möchte ich mich bei meinem Erstgutachter Prof. Dr. Ralf Münnich bedanken. Er hat mich im Rahmen meiner wissenschaftlichen Tätigkeit und der Anfertigung dieser Dissertation immer voll unterstützt und ermutigt. Seine konstruktiven Kommentare haben diese Arbeit bereichert.

Ich möchte mich zudem bei meinem Zweitgutachter PD Dr. Siegfried Gabler bedanken. Dies gilt insbesondere für seine Unterstützung und die belebenden Diskussionen, welche zum Gelingen dieser Arbeit beigetragen haben.

Weiterhin möchte ich meinen früheren Kollegen vom Lehrstuhl für Wirtschafts- und Sozialstatistik sowie meinen aktuellen Kollegen vom German Internet Panel (GIP) für die große Unterstützung und zahlreichen Diskussionen bedanken. Ein ganz großes Dankeschön gilt hier insbesondere Katharina Burgdorf, Tobias Enderle, Michael Englert, Florian Ertz, Franziska Gebhard, Laura Güdemann, Jessica Herzing, Hennig Höllwarth, Jan-Philipp Kolb, Hariolf Merkle, Martin Rupp, Julian Wagner, Thomas Zimmermann und Stefan Zins.

Ein ganz besonderer Dank gilt meinen Eltern Cornelia und Heinrich Bruch, sowie meinem Bruder Martin Bruch, denen ich diese Arbeit widme. Sie haben mich in den letzten Jahren im meinem Lebensweg immer unglaublich unterstützt.

Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Tabellenverzeichnis	V
Symbolverzeichnis	VI
1 Varianzschätzung unter Imputation und bei komplexen Stichproben- designs	1
1.1 Motivation und Thema der Arbeit	1
1.2 Aufbau der Arbeit	2
2 Einordnungen und Systematisierungen von fehlenden Werten und Imputationsmethoden	4
2.1 Arten von fehlenden Werten und dazugehörige Mechanismen	4
2.2 Imputation fehlender Werte	6
2.2.1 Klassifikation von Imputationsmethoden	7
2.2.2 Mittelwertimputation	9
2.2.3 Verhältnisimputation	10
2.2.4 Regressionsimputation	11
2.2.5 Hot-Deck-Random-Imputation	13
2.2.6 Nearest-Neighbour-Imputation	15
2.2.7 Predictive-Mean-Matching	16
2.2.8 Fractional-Imputation	17
2.2.9 Varianzen der imputierten Werte der verschiedenen Imputations- methoden	18
2.2.10 Zusammenfassende Gegenüberstellung der Imputationsmethoden . .	24
3 Varianzschätzung unter fehlenden Werten und Imputation	25
3.1 Varianzschätzung für Totalwertschätzer im Rahmen verschiedener Stich- probendesigns bei vollständigen Beobachtungen	26
3.2 Fälle der Entstehung von fehlenden Werten	29
3.3 Varianz bei fehlenden Werten	33
3.4 Varianz bei Imputation	35
3.4.1 Quellen der Unsicherheit und die dazugehörenden Varianzkompo- nenten bei Imputation hinsichtlich der Entstehung fehlender Werte	35
3.4.2 Formale Zerlegung der Gesamtvarianz in einzelne Varianzkompo- nenten	39
3.4.3 Beitrag der einzelnen Komponenten zur Gesamtvarianz	42
3.5 Varianz bei bestimmten Imputationsmethoden	44
3.6 Direkte Varianzschätzung	46
3.6.1 Mittelwertimputation	46
3.6.2 Hot-Deck-Random-Imputation	47
3.6.3 Verhältnisimputation	48

3.6.4	Regressionsimputation	49
3.6.5	Nearest-Neighbour-Imputation	50
3.6.6	Kritische Bewertung der direkten Varianzschätzer	51
4	Resampling Verfahren	52
4.1	Anwendung der Resampling-Verfahren zur Schätzung der Varianz	52
4.1.1	Allgemeine Vorgehensweise	52
4.1.2	Resampling-Verfahren bei fehlenden Werten und Imputation	53
4.1.3	Resampling-Verfahren bei mehrstufigen Stichprobendesigns	56
4.1.4	Vor- und Nachteile von Resampling-Verfahren	58
4.2	Random-Groups	59
4.3	Jackknife	60
4.3.1	Delete-a-Group-Jackknife	61
4.3.2	Jackknife bei fehlenden Werten und Imputation	62
4.4	Bootstrap	63
4.4.1	Monte-Carlo-Bootstrap bei fehlenden Werten und Imputation	64
4.4.2	Repeated-Half-Sample-Bootstrap	65
4.5	Rescaling-Bootstrap ohne Zurücklegen	66
4.5.1	Ursprüngliche Version nach Rao und Wu	66
4.5.2	Allgemeines Vorgehen	67
4.5.3	Rescaling-Bootstrap ohne Zurücklegen bei mehrstufigen Stichprobendesigns	68
4.5.4	Rescaling-Bootstrap ohne Zurücklegen bei fehlenden Werten und Imputation	71
4.5.5	Rescaling-Bootstrap bei fehlenden Werten für mehrstufige Stichprobendesigns	73
4.6	Balanced-Repeated-Replication	75
4.6.1	Allgemeine Beschreibung	75
4.6.2	Vergleich zwischen BRR und Rescaling-Bootstrap	76
4.6.3	BRR bei fehlenden Werten und Imputation	78
4.7	Resampling-Verfahren unter Berücksichtigung der Varianzkomponente bezüglich fehlender Werte	80
4.7.1	Allgemeine Problematik	80
4.7.2	Verfahren zur Berücksichtigung der Varianzkomponente bezüglich fehlender Werte	82
4.7.3	Berücksichtigung der Varianzkomponente bezüglich fehlender Werte über Korrekturfaktoren	83
5	Simulation	92
5.1	Simulationsbeschreibung	92
5.1.1	Einstufiges stratifiziertes Stichprobendesign	92
5.1.2	Mehrstufiges stratifiziertes Stichprobendesign	95
5.1.3	Varianzzerlegung	98
5.1.4	Erzeugung der fehlenden Werte	102
5.1.5	Szenarien	106
5.1.6	Schätzer	107
5.1.7	Benchmark Problematik	110
5.2	Ergebnisse bezüglich des einstufigen stratifizierten Stichprobendesigns	112
5.2.1	Vergleich verschiedener Varianzschätzmethoden	112
5.2.2	Varianzschätzung für komplexere Imputationsmethoden	121

5.2.3	Untersuchungen zu bestimmten Problematiken der Varianzschätzung unter Imputation	127
5.2.4	Große Auswahlssätze	132
5.3	Varianzzerlegung bezüglich der einfachen Zufallsstichprobe	136
5.3.1	Varianzzerlegung bezüglich der drei Fälle an fehlenden Werten und Mittelwertimputation	137
5.3.2	Varianzzerlegung bezüglich unterschiedlicher Auswahlssätze bei Mittelwertimputation	142
5.3.3	Varianzzerlegung hinsichtlich verschiedener Ausfallraten bei Mittelwertimputation	143
5.3.4	Varianzzerlegung hinsichtlich verschiedener Imputationsmethoden .	144
5.3.5	Varianzzerlegung bezüglich der drei Fälle an fehlenden Werten für Hot-Deck-Random-Imputation	145
5.3.6	Varianzzerlegung für verschiedene Parameter für Hot-Deck-Random-Imputation	148
5.3.7	Fazit bezüglich der Zerlegung der Gesamtvarianz	150
5.4	Ergebnisse bezüglich der mehrstufigen Stichprobendesigns	150
5.4.1	Kleine und mittlere Auswahlssätze	151
5.4.2	Große Auswahlssätze	157
5.4.3	Vorliegen einer ungleichen PSU-Größe	161
5.4.4	Varianzzerlegung im mehrstufigen Stichprobendesign ohne Stratifizierung	164
5.4.5	Fazit bezüglich der Behandlung mehrstufiger Stichprobendesigns . .	168
6	Handlungsempfehlungen und Ausblick	170
6.1	Wichtige Erkenntnisse und Handlungsempfehlungen	170
6.2	Ausblick	172
A	Anhang	173
A.1	Herleitung der Korrekturfaktoren aus Abschnitt 4.7.3	173
A.1.1	Herleitung des Korrekturfaktors für Mittelwertimputation bei der Entstehung variabler fehlender Werte in der Grundgesamtheit aus (4.49)	173
A.1.2	Herleitung des Korrekturfaktors für Hot-Deck-Random-Imputation bei Entstehung variabler fehlender Werte in der Grundgesamtheit aus (4.58)	174
A.1.3	Herleitung des Korrekturfaktors für Hot-Deck-Random-Imputation für den Fall eines fixen Response-Vektors in der Grundgesamtheit aus (4.64)	175
A.2	Simulationsbeschreibung	176
A.2.1	Methodenübersicht	176
A.2.2	Szenarien	183
A.2.3	Programm zur Erzeugung der Schichten und PSU in \mathbf{R} für das mehrstufige Stichprobendesign mit Stratifizierung auf der ersten Stufe	197
A.3	Simulationsergebnisse	198
A.3.1	Punktschätzer bezüglich des Vergleichs unterschiedlicher Ausfallmechanismen in Abschnitt 5.2.3.3	198
A.3.2	Vergleich der direkten Varianzschätzer im Falle von Regressionsimputation	199

A.3.3	Regressionsimputation mehrstufiges Stichprobendesign mit einer Stratifizierung auf der ersten Stufe	200
-------	---	-----

Abbildungsverzeichnis

2.1	Verteilungen der imputierten Werte der einzelnen Methoden in der Grundgesamtheit 1 bei geringer Korrelation	19
2.2	Verteilungen der imputierten Werte der einzelnen Methoden in der Grundgesamtheit 2 bei hoher Korrelation	20
2.3	Verteilungen des Untersuchungsmerkmals y nach der Imputation mit den verschiedenen Methoden in der Grundgesamtheit 1 bei geringer Korrelation	22
2.4	Verteilungen des Untersuchungsmerkmals y nach der Imputation mit den verschiedenen Methoden in der Grundgesamtheit 2 bei hoher Korrelation .	23
3.1	Fall 1: Entstehung variabler fehlender Werte in der Grundgesamtheit . . .	30
3.2	Fall 2: Entstehung fehlender Werte in der Stichprobe	31
3.3	Fall 3: Einmalig generierter (fixer) Response-Vektor in der Grundgesamtheit	32
3.4	Quellen der Unsicherheit mit dazugehörigen Varianzkomponenten in Fall 1 und der Verwendung eines zufälligen Imputationsverfahrens	36
3.5	Quellen der Unsicherheit mit dazugehörigen Varianzkomponenten in Fall 1 und der Verwendung eines deterministischen Imputationsverfahrens . . .	37
3.6	Quellen der Unsicherheit mit dazugehörigen Varianzkomponenten in Fall 2 und der Verwendung eines zufälligen Imputationsverfahrens	38
3.7	Quellen der Unsicherheit mit dazugehörigen Varianzkomponenten in Fall 3 und der Verwendung eines zufälligen Imputationsverfahrens	39
5.1	Histogramm und Kerndichteschätzung des Untersuchungsmerkmals im einstufigen stratifizierten Stichprobendesign	93
5.2	Histogramm und Kerndichteschätzung der metrischen Hilfsvariable im einstufigen stratifizierten Stichprobendesign	94
5.3	Histogramm und Kerndichteschätzung des Untersuchungsmerkmals im Szenario des mehrstufigen Stichprobendesigns	97
5.4	Verteilung fehlender Werte für die beiden Hilfsvariablen im einstufigen stratifizierten Design	104
5.5	Verteilung fehlender Werte für die beiden Hilfsvariablen im mehrstufigen Design mit einer Stratifizierung auf der ersten Stufe	105
5.6	Stabilität der Monte-Carlo-Varianz für den Fall von Mittelwertimputation, kleinen Auswahlsätzen und der Entstehung variabler fehlender Werte in der Grundgesamtheit	110
5.7	Stabilität der Monte-Carlo-Varianz für den Fall von Mittelwertimputation, kleinen Auswahlsätzen und der Entstehung fehlender Werte in der Stichprobe	111
5.8	Vergleich der Varianzschätzer im Fall von Mittelwertimputation, kleiner Auswahlsätze, Entstehung variabler fehlender Werte in der Grundgesamtheit und MAR	113

5.9	Vergleich des Rescaling-Bootstrap ohne Zurücklegen für die drei Fälle an fehlenden Werten bei kleinen Auswahlstäben, MAR und Mittelwertimputation	115
5.10	Vergleich der Varianzschätzer im Fall von Verhältnisimputation, kleiner Auswahlstabe, Entstehung variabler fehlender Werte in der Grundgesamtheit und MAR	116
5.11	Vergleich der Varianzschätzer im Fall von Regressionsimputation, kleiner Auswahlstabe, Entstehung variabler fehlender Werte in der Grundgesamtheit und MAR	118
5.12	Vergleich der Varianzschätzer im Fall von Hot-Deck-Random-Imputation, kleiner Auswahlstabe, Entstehung variabler fehlender Werte in der Grundgesamtheit und MAR	119
5.13	Vergleich der Reimputation und Adjustierung für den Rescaling-Bootstrap ohne Zurücklegen im Rahmen von Hot-Deck-Random-Imputation	120
5.14	Vergleich der Varianzschätzer im Fall der stochastischen Regressionsimputation, der Entstehung variabler fehlender Werte in der Grundgesamtheit, kleiner Auswahlstabe und MAR	122
5.15	Vergleich der Varianzschätzer im Fall von Nearest-Neighbour-Imputation, der Entstehung variabler fehlender Werte in der Grundgesamtheit, kleiner Auswahlstabe und MAR	123
5.16	Vergleich der Varianzschätzer im Fall von Nearest-Neighbour-Imputation bei Schätzung der Gesamtanzahl der Kategorie 2 der Haushaltsgröße und des Einkommens als Hilfsvariable	124
5.17	Varianzschätzung im Fall von Predictive-Mean-Matching, der Entstehung variabler fehlender Werte in der Grundgesamtheit, kleiner Auswahlstabe und MAR	125
5.18	Vergleich der Varianzschätzer im Fall von Fractional-Imputation, der Entstehung variabler fehlender Werte in der Grundgesamtheit, kleiner Auswahlstabe und MAR	126
5.19	Erhöhung der Ausfallrate bei Mittelwertimputation, Erzeugung variabler fehlender Werte in der Grundgesamtheit, kleinen Auswahlstabe und MAR	128
5.20	Erzeugung variabler fehlender Werte in der Grundgesamtheit in den gering besetzten Kategorien bei Mittelwertimputation, kleinen Auswahlstabe und MAR	129
5.21	Vergleich der verschiedenen Ausfallmechanismen für Mittelwertimputation, die Entstehung variabler fehlender Werte in der Grundgesamtheit und kleinen Auswahlstabe	130
5.22	Kombination aus Mittelwert- und Hot-Deck-Random-Imputation bei kleinen Auswahlstäben und Entstehung variabler fehlender Werten in der Grundgesamtheit	131
5.23	Vergleich der Varianzschätzer im Fall großer Auswahlstäbe für Mittelwertimputation und MAR	133
5.24	Vergleich der Varianzschätzer im Fall großer Auswahlstäbe für Hot-Deck-Random-Imputation und MAR	135
5.25	Varianzzerlegung hinsichtlich der drei Fälle an fehlenden Werten für Mittelwertimputation, mittlerem Auswahlstabe und mittlerer Rate an fehlenden Werten	137

5.26	Abdeckung der einzelnen Varianzkomponenten durch den Rescaling-Bootstrap für einen mittleren Auswahlsatz, Mittelwertimputation und der Entstehung variabler fehlender Werte in der Grundgesamtheit	138
5.27	Abdeckung der einzelnen Varianzkomponenten durch den Rescaling-Bootstrap für einen mittleren Auswahlsatz, Mittelwertimputation und der Entstehung fehlender Werte in der Stichprobe	140
5.28	Abdeckung der einzelnen Varianzkomponenten durch den Rescaling-Bootstrap für einen mittleren Auswahlsatz, Mittelwertimputation und Erzeugung eines fixen Response-Vektors in der Grundgesamtheit	141
5.29	Varianzzerlegung hinsichtlich verschiedener Auswahlsätze für Mittelwertimputation, Entstehung variabler fehlender Werte in der Grundgesamtheit und mittlerer Rate an fehlenden Werten	142
5.30	Varianzzerlegung hinsichtlich verschiedener Ausfallraten für Mittelwertimputation, Entstehung variabler fehlender Werte in der Grundgesamtheit und mittlerem Auswahlsatz	143
5.31	Varianzzerlegung hinsichtlich verschiedener Imputationsmethoden bei Entstehung variabler fehlender Werte in der Grundgesamtheit, mittlerer Ausfallrate und mittlerem Auswahlsatz	144
5.32	Varianzzerlegung hinsichtlich der drei Fälle an fehlenden Werten für Hot-Deck-Random-Imputation, mittlerer Ausfallrate und mittlerem Auswahlsatz	145
5.33	Abdeckung der einzelnen Varianzkomponenten durch den Rescaling-Bootstrap für einen mittleren Auswahlsatz, Hot-Deck-Random-Imputation und Entstehung variabler fehlender Werte in der Grundgesamtheit	146
5.34	Abdeckung der einzelnen Varianzkomponenten durch den Rescaling-Bootstrap für einen mittleren Auswahlsatz, Hot-Deck-Random-Imputation und eines fixen Response-Vektors in der Grundgesamtheit	147
5.35	Varianzzerlegung hinsichtlich verschiedener Auswahlsätze für Hot-Deck-Random-Imputation, Entstehung variabler fehlender Werte in der Grundgesamtheit und mittlerer Auswahlrate	148
5.36	Varianzzerlegung hinsichtlich verschiedener Ausfallraten für Hot-Deck-Random-Imputation, Entstehung variabler fehlender Werte in der Grundgesamtheit und mittlerem Auswahlsatz	149
5.37	Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, kleinen Auswahlsätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit	152
5.38	Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, mittleren Auswahlsätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit	153
5.39	Vergleich beim zweistufigen Stichprobendesign, Hot-Deck-Random-Imputation, kleinen Auswahlsätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit .	155
5.40	Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, kleinen Auswahlsätzen, homogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit	156
5.41	Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, hohen Auswahlsätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit	158

5.42	Vergleich beim zweistufigen Stichprobendesign, Hot-Deck-Random-Imputation, großen Auswahl­sätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit	159
5.43	Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, großen Auswahl­sätzen, homogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit	160
5.44	Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, kleinen Auswahl­sätzen, heterogener PSU, einer gering variierenden PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit	161
5.45	Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, kleinen Auswahl­sätzen, heterogener PSU, einer stark variierenden PSU-Größe, MAR und Entstehung variabler fehlenden Werte in der Grundgesamtheit	162
5.46	Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, großen Auswahl­sätzen, heterogener PSU, einer stark variierenden PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit	163
5.47	Varianzzerlegung für den Fall kleiner Auswahl­sätze, heterogener PSU und einer gleichen PSU-Größe	165
5.48	Varianzzerlegung für den Fall kleiner Auswahl­sätze, homogener PSU und einer gleichen PSU-Größe	166
5.49	Varianzzerlegung für den Fall kleiner Auswahl­sätze, heterogener PSU und einer ungleichen PSU-Größe	167
5.50	Varianzzerlegung für den Fall großer Auswahl­sätze, heterogener PSU und einer ungleichen PSU-Größe	168
A.1	Vergleich der Punktschätzer unter unterschiedlichen Ausfallmechanismen	198
A.2	Vergleich der direkten Varianzschätzer im Falle von Regressionsimputation und der einfachen Zufallsstichprobe	199
A.3	Vergleich beim zweistufigen Stichprobendesign, Regressionsimputation, kleinen Auswahl­sätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit	200
A.4	Vergleich beim zweistufigen Stichprobendesign, Regressionsimputation, großen Auswahl­sätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit	201

Tabellenverzeichnis

2.1	Einteilung der Imputationsverfahren	8
2.2	Bandbreite der Imputationswerte	21
2.3	Gegenüberstellung der Imputationsmethoden	24
4.1	Adjustierung imputierter Werte	55
5.1	Schichtgrößen	94
5.2	Größe der einzelnen Kategorien	95
5.3	Größe der einzelnen Kategorien im mehrstufigen Design	98
5.4	Notwendigkeit der Berücksichtigung von Korrekturfaktoren (KF)	136

Symbolverzeichnis

Symbol	Beschreibung
a	Substichprobe a
b	Substichprobe b
α	Parameter
$\tilde{\alpha}$	Parameter bei der Erzeugung fehlender Werte
$\alpha_{HotDeck}$	Korrekturfaktor für Hot-Deck-Random-Imputation
$\hat{\alpha}_{HotDeck}$	Geschätzter Korrekturfaktor für Hot-Deck-Random-Imputation
α_I	Korrekturfaktor für eine bestimmte Imputationsmethode
$\hat{\alpha}_I$	Geschätzter Korrekturfaktor für eine bestimmte Imputationsmethode
α_{MW}	Korrekturfaktor für Mittelwertimputation
$\hat{\alpha}_{MW}$	Geschätzter Korrekturfaktor für Mittelwertimputation
α_{Reg}	Korrekturfaktor für Regressionsimputation
$\hat{\alpha}_{Reg}$	Geschätzter Korrekturfaktor für Regressionsimputation
β	Koeffizient
$\tilde{\beta}$	Parameter bei der Erzeugung fehlender Werte
Cor	Korrelationskoeffizient
Cov	Kovarianz
Cov_p	Kovarianz bezüglich der Stichprobenverteilung
d_i	Anzahl der Verwendung von Einheit i als Spender
D	Anzahl der Spender für einen fehlenden Wert
δ_i	Indikator ob ein Element i in die Substichprobe gelangt mit Ausprägung 1 oder 0
δ_{hi}	Indikator ob ein Element i in Schicht h in die Substichprobe gelangt mit Ausprägung 1 oder 0
δ_{hd}	Indikator ob eine PSU d in Schicht h in die Substichprobe gelangt mit Ausprägung 1 oder 0
δ_{hdi}	Indikator ob eine Einheit i in PSU d in Schicht h in die Substichprobe gelangt mit Ausprägung 1 oder 0
E_p	Erwartungswert bezüglich der Stichprobenverteilung
E_q	Erwartungswert bezüglich der Response-Verteilung
E_I	Erwartungswert bezüglich der Imputationsmechanismus
f	Gesamtauswahlsatz
f_h	Auswahlsatz in einer bestimmten Schicht h
f_{hd}	Auswahlsatz in einer PSU d einer bestimmten Schicht h
$\tilde{\gamma}$	Parameter bei der Erzeugung fehlender Werte
H	Anzahl der Schichten
K	Anzahl an Hilfsvariablen
κ	Kategorie, in der die Imputation stattfindet, z.B. Imputationsklasse oder Schicht
$K1$	Gebildete Kategorie beim mehrstufigen Rescaling-Bootstrap ohne Zurücklegen für Einheiten mit $\delta_{hd} = 1$ und $\delta_{hdi} = 0$

Symbol	Beschreibung
$K2$	Gebildete Kategorie beim mehrstufigen Rescaling-Bootstrap ohne Zurücklegen für Einheiten mit $\delta_{hd} = 0$ und $\delta_{hdi} = 1$
$K3$	Gebildete Kategorie beim mehrstufigen Rescaling-Bootstrap ohne Zurücklegen für Einheiten mit $\delta_{hd} = 1$ und $\delta_{hdi} = 1$
$K4$	Gebildete Kategorie beim mehrstufigen Rescaling-Bootstrap ohne Zurücklegen für Einheiten mit $\delta_{hd} = 0$ und $\delta_{hdi} = 0$
Λ	Anzahl an Variablen im Datensatz
l	Anzahl der PSU in der Stichprobe
l^*	Anzahl der PSU in der Substichprobe
l_h	Anzahl der PSU in der Stichprobe in Schicht h
l_h^*	Anzahl der PSU in der Substichprobe in Schicht h
L	Anzahl der PSU in der Grundgesamtheit
L_h	Anzahl der PSU in der Grundgesamtheit in Schicht h
m	Anzahl der fehlenden Einheiten
M	Menge der fehlenden Einheiten
M^*	Menge der fehlenden Einheiten in der Substichprobe
n	Umfang der gesamten Stichprobe
n_d	Umfang der Stichprobe in PSU d
n_h	Umfang der Stichprobe in Schicht h
n_{hd}	Umfang der Stichprobe in PSU d in Schicht h
n^*	Umfang der Substichprobe für ein bestimmtes Resampling-Verfahren
n_h^*	Umfang der Substichprobe für ein bestimmtes Resampling-Verfahren in Schicht h
N	Größe der Grundgesamtheit
N_h	Größe der Grundgesamtheit in Schicht h
N_d	Größe der Grundgesamtheit in PSU d
Ω	Anzahl der gezogenen Substichproben
p	Wahrscheinlichkeit
p_R	Antwort-/ Teilnahmewahrscheinlichkeit
p_{Ri}	Antwort-/ Teilnahmewahrscheinlichkeit der Einheit i
ϕ	Unbekannter Parameter
Φ	Platzhalter bedingte Größe
q	Response-Verteilung
r	Anzahl der antwortenden Einheiten
r_h	Anzahl der antwortenden Einheiten in der Schicht h
R	Menge der antwortenden Einheiten
R^*	Menge der antwortenden Einheiten in der Substichprobe
s	Stichprobenvektor
s^2	Geschätzte Varianz

Symbol	Beschreibung
σ^2	Varianz des Untersuchungsmerkmals in der Grundgesamtheit
σ_u^2	Varianz der Störgröße in der Grundgesamtheit
T	Funktion zur Berechnung der interessierenden Statistik
τ	Totalwert
$\hat{\tau}$	Geschätzte Totalwert
θ	Allgemeine Statistik
$\hat{\theta}$	Allgemeiner Schätzer
$\hat{\theta}_I$	Imputierter Schätzer
$\hat{\theta}_{NR}$	Schätzer unter fehlenden Werten
u	Störterm
$V(\cdot)$	Varianz
V_p	Varianz bezüglich der Stichprobenverteilung
V_q	Varianz bezüglich der Response-Verteilung
V_I	Varianz bezüglich des Imputationsmechanismus
V_1	Varianz bei vollständigen Beobachtungen auf der ersten Stufe des Stichprobendesigns
V_{NR}^*	Varianzkomponente bezüglich fehlender Werte eines bestimmten Falls * bei einem deterministischen Imputationsverfahren
V_{NR2}^1	Varianzkomponente bezüglich fehlender Werte hinsichtlich des Falls der Entstehung variabler fehlender Werte in der Grundgesamtheit bei einem deterministischen Imputationsverfahren auf der zweiten Stufe des Stichprobendesigns
V_S^*	Stichprobenkomponente der Gesamtvarianz bezüglich eines bestimmten Falls * bei einem deterministischen Imputationsverfahren
V_{S1}^1	Stichprobenkomponente der Gesamtvarianz bezüglich des Falls der Entstehung variabler fehlender Werte in der Grundgesamtheit bei einem deterministischen Imputationsverfahren auf der ersten Stufe des Stichprobendesigns
V_{S2}^1	Stichprobenkomponente der Gesamtvarianz bezüglich des Falls der Entstehung variabler fehlender Werte in der Grundgesamtheit bei einem deterministischen Imputationsverfahren auf der zweiten Stufe des Stichprobendesigns
\tilde{V}_{Imp}^*	Imputationskomponente der Gesamtvarianz bezüglich eines bestimmten Falls * bei einem zufälligen Imputationsverfahren
\tilde{V}_{NR}^*	Varianzkomponente bezüglich fehlender Werte eines bestimmten Falls * bei einem zufälligen Imputationsverfahren
\tilde{V}_S^*	Stichprobenkomponente der Gesamtvarianz bezüglich eines bestimmten Falls * bei einem zufälligen Imputationsverfahren

Symbol	Beschreibung
w_i	Designgewicht der Einheit i
w_i^*	Nach einem Resampling-Verfahren angepasster Gewichtevektor einer Einheit in der Substichprobe
w^*	Nach einem Resampling-Verfahren angepasster Gewichtevektor der Einheiten in der Substichprobe
w_{-1}	Nach dem Delete-1-Jackknife angepasster Gewichtevektor der Einheiten in der Substichprobe
w_{boot}^*	Die Designgewichte der ursprünglichen Stichprobe der in die Bootstrap Stichprobe gelangten Elemente beim Monte-Carlo-Bootstrap
w_{BRR}	Nach dem BRR angepasster Gewichtevektor aller Einheiten der Stichprobe
w_h	Designgewichte im Bezug auf Schicht h
w_{hd}	Designgewichte im Bezug auf Schicht h und PSU d
w_{hdi}	Designgewichte im Bezug auf Schicht h und PSU d und Einheit i
w_{hi}	Designgewichte im Bezug Einheit i in Schicht h
$w_{h,-1}$	Nach dem geschichteten Delete-1-Jackknife angepasster Gewichtevektor der Einheiten in der Substichprobe
$w_{h,-g}$	Nach dem geschichteten Delete-a-Group-Jackknife angepasster Gewichtevektor der Einheiten in der Substichprobe
w_g	Nach der Random-Group-Methode angepasster Gewichtevektor der Einheiten in der Substichprobe
w_{-g}	Nach dem Delete-a-Group-Jackknife angepasster Gewichtevektor der Einheiten in der Substichprobe
w_{repeat}^*	Die Designgewichte der ursprünglichen Stichprobe der beim Repeated-Half-Sample-Bootstrap in die Substichprobe gelangten Elemente
w_{Resc}^*	Nach dem Rescaling-Bootstrap ohne Zurücklegen angepasster Gewichtevektor aller Einheiten der Stichprobe
\tilde{w}_{ij}	Fraction
Ξ	Sonstige Bedingungen und Variablen
x	Hilfsvariable allgemein
x_k	Hilfsvariable k
y_i	Ausprägung des Untersuchungsmerkmals bei Einheit i
y	Untersuchungsmerkmal
y_S	Untersuchungsmerkmal bezüglich der in die Stichprobe gelangenden Einheiten
y^*	Für Substichprobe gezogene y

Symbol	Beschreibung
$y^{*,adj}$	Adjustierter Vektor des Untersuchungsmerkmals der Stichprobe
$y^{*,reimp}$	Reimputierter Vektor des Untersuchungsmerkmals der Stichprobe
\tilde{y}_j	Imputationswert für die fehlende Einheit j
\bar{y}	Mittelwert
\bar{y}_U	Mittelwert in der Grundgesamtheit
$\hat{\bar{y}}$	Geschätzter Mittelwert
\hat{y}_i	Geschätzter/prognostizierter Wert der Einheit i aus dem Regressionsmodell
z	Response-Vektor
Z	Response-Matrix
ζ	Datensatz
$\dot{\cup}$	Disjunkte Vereinigungsmenge

1 Varianzschätzung unter Imputation und bei komplexen Stichprobendesigns

1.1 Motivation und Thema der Arbeit

Um Aussagen über die Qualität der Schätzung einer interessierenden Statistik zu gewinnen ist die Schätzung der Varianz des entsprechenden Punktschätzers von entscheidender Bedeutung. So basieren Qualitätsmaße wie z.B. Konfidenzintervalle oder der Root-Mean-Squared-Error auf einer validen Varianzschätzung. In der Praxis und hier insbesondere für die Amtliche Statistik stellen sich in diesem Zusammenhang insbesondere zwei Herausforderungen. Zum einen zeichnen sich viele Erhebungen durch das Auftreten von fehlenden Werten aus. Dies ist zum Beispiel im Rahmen von Haushaltserhebungen häufig im Zusammenhang mit Variablen wie dem Einkommen der Fall. Dadurch können in Abhängigkeit des Ausfallmechanismus verzerrte Punktschätzer resultieren. Zur Kompensation der fehlenden Werte stehen jedoch verschiedene Methoden zur Verfügung. Im Rahmen dieser Arbeit wird die Imputation in den Mittelpunkt gestellt. Zum anderen besitzen Stichprobendesigns in der Praxis meist eine hohe Komplexität. Als Beispiele sind Stratifizierungen und mehrstufige Stichprobendesigns zu nennen, wobei bei letzteren die Komplexität mit größerer Anzahl an Stufen zunimmt. Sowohl fehlende Werte und die Durchführung der Imputation, als auch komplexe Stichprobendesigns haben große Auswirkungen auf die Varianz eines Punktschätzers. Es müssen geeignete Varianzschätzer gefunden werden. Diese Thematik steht im Mittelpunkt dieser Arbeit.

Die Motivation einer möglichst exakten Schätzung der Varianz ergibt sich aus dem Ziel einer validen Qualitätsmessung der durchgeführten Inferenz. Werden bestimmte Komponenten der Varianz durch die Schätzung nicht erfasst, kann die Varianzberechnung zu niedrig sein. Die Folge wäre, dass z.B. Kenngrößen wie der Root-Mean-Squared-Error zu gering ausgewiesen werden oder die Breite von Konfidenzintervallen, welche für Aussagen über die Genauigkeit eines Schätzers von Bedeutung ist, unterschätzt wird. Bei einer Überschätzung der Varianz gelten die gegenteiligen Aussagen. Folglich können Überschätzungen wie auch Unterschätzungen zu erheblichen Fehlschlüssen bei der Interpretation bezüglich der Genauigkeit des Punktschätzers führen. Der Root-Mean-Squared-Error oder die Breite von Konfidenzintervallen sind nur zwei ausgewählte Beispiele. Weitere Beispiele sind die Durchführung von Hypothesentests oder die Berechnung von Designeffekten. Auch hier gehen Varianzschätzungen ein und Verzerrungen können erhebliche Fehlinterpretation mit sich bringen.

In der Literatur existiert eine Vielzahl von Möglichkeiten, wie eine Varianzschätzung im Falle fehlender Werte und der Anwendung von Imputation durchgeführt werden kann. In diesem Zusammenhang können z.B. die sogenannten Resampling-Verfahren verwendet werden. Jedoch beziehen sich hier die Arbeiten mit Bezug auf fehlende Werte und Imputation häufig auf klassische Ansätze wie den Monte-Carlo-Bootstrap oder den Delete-1-Jackknife. Neuere Verfahren wie z.B. der Rescaling-Bootstrap ohne Zurücklegen von

CHIPPERFIELD und PRESTON (2007) basieren hingegen auf dem Fall vollständiger Beobachtungen. Diese Methode ist gerade im Zusammenhang mit komplexen mehrstufigen Stichprobendesigns von entscheidender Bedeutung. Daher muss auch dieses Verfahren auf den Fall fehlender Werte und Anwendung von Imputation übertragen werden. Und dies stellt ein wichtiges Ziel dieser Arbeit dar. Es werden sowohl für ein- wie mehrstufige Stichprobendesigns Ansätze entwickelt, wie solche neueren Resampling-Verfahren bei Vorliegen von fehlenden Werten und deren Kompensation über Imputation angewendet werden können.

Des Weiteren findet in der Literatur häufig eine sogenannte Varianzzerlegung statt. Die Gesamtvarianz unter fehlenden Werten und Imputation wird hierbei in ihre einzelnen Komponenten zerlegt. Diese Bestandteile sind bezüglich ihrer Höhe und in ihrem Beitrag zur Gesamtvarianz von bestimmten Parametern abhängig. Dies ist für Resampling-Verfahren besonders wichtig, da diese Methoden in ihrer ursprünglichen Anwendung in der Lage sind, nur bestimmte Komponenten der Gesamtvarianz abzudecken (vgl. HAZIZA, 2010, MASHREGHI et al., 2014). Daher gilt es, die Parameterkonstellationen zu erörtern, bei welchen die Komponenten der Gesamtvarianz, welche durch die Resampling-Verfahren nicht erfasst werden können, von Bedeutung sind. Für solche Situationen müssen Ansätze verwendet werden, welche auf die Berücksichtigung dieser Varianzkomponenten abstellen. Solche Verfahren werden zum Beispiel in MASHREGHI et al. (2014) gegeben, welche auch im Rahmen dieser Arbeit von Interesse sind.

Für die Varianzzerlegung ist insbesondere auch entscheidend, an welcher Stelle des Erhebungsprozesses fehlende Werte entstehen und welche Annahmen bezüglich der fehlenden Werte getroffen werden. Im Rahmen dieser Arbeit werden drei Fälle der Entstehung fehlender Werte berücksichtigt, welche ausführlich erläutert werden und unterschiedliche Konsequenzen bezüglich der Varianz und ihrer Komponenten mit sich bringen.

Zur Überprüfung der verschiedenen Varianzschätzmethoden hinsichtlich ihrer Eignung zur Schätzung der Varianz unter fehlenden Werten und Anwendung von Imputation bei komplexen Stichprobendesigns wird eine Monte-Carlo-Simulationsstudie durchgeführt. In Abhängigkeit des untersuchten Stichprobendesigns werden verschiedene Grundgesamtheiten verwendet. Ebenfalls wird mit Hilfe einer Monte-Carlo-Simulation in dieser Arbeit eine Zerlegung der Varianz in ihre einzelnen Komponenten vorgenommen. Durch dieses Vorgehen ist es möglich, eine Varianz auch für komplexere Situationen zu zerlegen. So wird in der vorliegenden Arbeit auch für den Fall eines mehrstufigen Stichprobendesigns bei Vorliegen von fehlenden Werten und Imputation eine Varianzzerlegung vorgenommen.

1.2 Aufbau der Arbeit

Der Arbeit liegt der folgende Aufbau zugrunde. In Kapitel 2 werden die Grundlagen zu fehlenden Werten und Imputationsmethoden behandelt. Im Abschnitt 2.1 werden zunächst verschiedene Arten fehlender Werte erörtert. Hier sind insbesondere die verschiedenen Antwort- bzw. Ausfallmechanismen zur Erzeugung fehlender Werte von Interesse, da diese für die Annahmen der verschiedenen Varianzschätzmethoden relevant sind. Anschließend erfolgt in Abschnitt 2.2 die Vorstellung der in dieser Arbeit verwendeten Imputationsmethoden. Es werden die Vor- und Nachteile ihrer Anwendung erörtert, wobei insbesondere ihre Auswirkungen auf die resultierenden Imputationswerte von Bedeutung sind.

In Kapitel 3 wird auf das zentrale Thema dieser Arbeit, die Varianzschätzung unter fehlenden Werten und deren Kompensation über Imputation, eingegangen. In Abschnitt 3.1 werden zunächst die Varianzen und ihre Schätzer für den Fall vollständiger Beobachtungen beschrieben. Anschließend erfolgt in Abschnitt 3.2 die Erläuterung der Fälle der Ent-

stehung fehlender Werte. Diese Thematik ist in Zusammenhang mit der durchgeführten Varianzzerlegung wichtig und wird im späteren Verlauf der Arbeit für die Definition der einzelnen Varianzkomponenten bezüglich der Stichprobenziehung, fehlender Werte und der Imputation benötigt. In Abschnitt 3.3 werden die allgemeinen Auswirkungen von fehlenden Werten auf die Varianz eines Schätzers einer interessierenden Statistik beschrieben. Die Effekte der Imputation als Maßnahme der Kompensation fehlender Werte auf die Varianz eines Schätzers werden in Abschnitt 3.4 gegeben. Hier wird insbesondere auch eine Varianzzerlegung nach drei verschiedenen Fällen der Entstehung fehlender Werte durchgeführt und der Beitrag der einzelnen Komponenten der Varianz in Abhängigkeit bestimmter Parameter beschrieben. In Abschnitt 3.5 werden konkrete Beispiele für die Varianz eines Schätzers bezüglich einzelner Imputationsmethoden gegeben. Der letzte Abschnitt des Kapitels 3 beschäftigt sich mit der Darstellung direkter Schätzer für die Varianz bei Auftreten fehlender Werte und deren Kompensation über Imputation.

Die Anwendung direkter Varianzschätzer im Falle fehlender Werte und Imputation basieren häufig auf strengen Annahmen und Modellen, deren Nichteinhaltung Verzerrungen mit sich bringen können. Um diesen Nachteilen zu begegnen, können Resampling-Verfahren angewendet werden, welche in dieser Arbeit von zentraler Bedeutung sind. Daher wird diesen Methoden ein eigenes Kapitel gewidmet. In Kapitel 4 werden diese Verfahren vorgestellt. Sie werden bezüglich ihrer Anwendung im Falle fehlender Werte und deren Kompensation über Imputation, sowie bezüglich ihrer Anwendung bei komplexen Stichprobendesigns erläutert. Es werden nacheinander verschiedene Resampling-Verfahren erörtert. Für neuere Resampling-Verfahren wie den Rescaling-Bootstrap ohne Zurücklegen werden in dieser Arbeit Ansätze vorgestellt, wie sie auf den Fall des Auftretens von fehlenden Werten und der Durchführung von Imputation übertragen werden können. Im Rahmen der Abschnitte 4.1 bis 4.6 werden bestimmte Annahmen zur Vernachlässigung bestimmter Varianzkomponenten getroffen. Diese werden in Abschnitt 4.7 aufgelöst. Hier werden Ansätze beschrieben, wie die zuvor vernachlässigten Varianzkomponenten erfasst werden können.

Zur Überprüfung der Eignung der vorgestellten Varianzschätzmethode zur Schätzung der Varianz eines Punktschätzers bei Auftreten fehlender Werte und Durchführung der Imputation bei komplexen Stichprobendesigns, wird in Kapitel 5 eine Monte-Carlo-Simulation durchgeführt. Hier erfolgt zunächst in Abschnitt 5.1 eine ausführliche Beschreibung dieser Simulation. Anschließend werden die Ergebnisse interpretiert. Begonnen wird in Abschnitt 5.2 mit den Resultaten aus dem einstufigen stratifizierten Stichprobendesign. Es folgt in Abschnitt 5.3 die Durchführung der Varianzzerlegung im Falle der einstufigen Zufallsstichprobe mit Hilfe der Monte-Carlo-Simulation. In Abschnitt 5.4 werden die Ergebnisse für die mehrstufigen Stichprobendesigns interpretiert.

Die Arbeit schließt mit einem Fazit. Hier werden die wichtigsten Aussagen der Arbeit aufgearbeitet, Handlungsempfehlungen gegeben und Ansatzpunkte zukünftiger Forschung eruiert.

2 Einordnungen und Systematisierungen von fehlenden Werten und Imputationsmethoden

2.1 Arten von fehlenden Werten und dazugehörige Mechanismen

Wie in dem vorherigen Kapitel beschrieben, können fehlende Werte große Auswirkungen auf die Punktschätzung und Varianzschätzung besitzen. Wichtig ist in diesem Zusammenhang, welche Art von fehlenden Werten vorliegt. Diese bringen verschiedene Problematiken mit sich. In der Literatur existieren unterschiedliche Möglichkeiten die Arten von fehlenden Werten zu klassifizieren. Ziel dieses Abschnittes ist es, einen Überblick über die gegebenen Klassifikationen und Definitionen zu geben.

Grundsätzlich wird angenommen, dass eine Menge von antwortenden Einheiten R durch eine unbekannte Wahrscheinlichkeitsverteilung q realisiert wird, welche Antwortmechanismus (bzw. Ausfallmechanismus bei gegenteiliger Betrachtung) genannt wird. Der Ort der Entstehung fehlender Werte kann ebenfalls Einfluss haben, wird jedoch zunächst nicht berücksichtigt. Diese Thematik wird in Abschnitt 3.2 aufgegriffen. In diesem Fall kann die Realisation fehlender Werte auf Basis der zugrunde liegenden Wahrscheinlichkeit p_R beschrieben werden durch: $q(R) = q(R|\Xi, y)$. Damit kann die Teilnahme z.B. vom Untersuchungsmerkmal y oder sonstigen Variablen und Bedingungen Ξ abhängig sein. Der Antwortmechanismus wird z.B. als *unconfounded* bezeichnet, wenn gilt: $q(R) = q(R|\Xi)$ mit einer Antwortwahrscheinlichkeit $p_R > 0$. Dies ist z.B. der Fall, wenn nur eine Abhängigkeit zu den Hilfsvariablen x_k für $k = 1, \dots, K$ besteht und nicht zum Untersuchungsmerkmal y . Trifft dies nicht zu, wird der Antwortmechanismus *confounded* genannt. Antworten die verschiedenen Einheiten unabhängig voneinander und besitzen sie gleiche Antwortwahrscheinlichkeiten $p_R > 0$, wird ein *unconfounded* Antwortmechanismus zudem als *uniform* definiert (vgl. LEE et al., 1994, S. 232). Wie noch zu erörtern sein wird, spielen uniforme Antwortmechanismen für die Punkt- und Varianzschätzung eine entscheidende Rolle. Eine Grundgesamtheit, bei der die Personen gleiche Antwortwahrscheinlichkeiten besitzen, ist nicht besonders realistisch. In der Praxis wird daher häufig versucht, Imputationsklassen über kategoriale Variablen zu bilden, in welchen die Einheiten mit gleichen Wahrscheinlichkeiten und unabhängig voneinander antworten (vgl. z.B. SHAO und SITTER, 1996, Abschnitt 2.4).

Die wohl bekannteste Einteilung von Mechanismen, welche zu beobachteten und fehlenden Werten führen, erfolgt nach Missing-Completely-at-Random (MCAR), Missing-at-Random (MAR) und Not-Missing-at-Random (NMAR). Diese drei Mechanismen werden ausführlich in LITTLE und RUBIN (2002, S. 12) beschrieben. Ausgangspunkt ist ein Datensatz ζ bestehend aus den Variablen ζ_k ($k = 1 \dots \Lambda$). Der Datensatz könnte z.B. aus dem interessierenden Untersuchungsmerkmal y und den Hilfsvariablen x_k für $k = 1, \dots, K$ bestehen. Weiterhin stellt ζ_{beob} die beobachteten Komponenten bzw. Einträge des Daten-

satzes und ζ_{fehl} die fehlenden Komponenten des Datensatzes dar. Die Matrix $Z = Z_{ki}$ nimmt den Wert 1 an, wenn ein Wert für die Einheit i bei der Variable k beobachtet wurde, und 0, wenn es sich um einen fehlenden Wert handelt. Somit stellt Z eine Matrix dar, welche die Indikatoren für beobachtete und fehlende Werte enthält. Der Ausfallmechanismus wird beschrieben durch die bedingte Verteilung von Z gegeben ζ und damit $q(Z|\zeta, \phi)$, wobei ϕ für die unbekannt Parameter steht. Von MCAR wird dann gesprochen, wenn die Teilnahme bzw. Nichtteilnahme, beschrieben durch Z , weder von den beobachtbaren Komponenten ζ_{beob} noch von den fehlenden Komponenten ζ_{fehl} abhängt. Damit gilt:

$$q(Z|\zeta, \phi) = q(Z|\phi), \text{ für alle } \zeta, \phi. \quad (2.1)$$

Ist der Ausfall hingegen von den beobachtbaren Komponenten ζ_{beob} , nicht aber von den fehlenden Komponenten ζ_{fehl} abhängig, so wird von MAR gesprochen. Es gilt:

$$q(Z|\zeta, \phi) = q(Z|\zeta_{beob}, \phi), \text{ für alle } \zeta_{fehl}, \phi. \quad (2.2)$$

Ein Ausfallmechanismus NMAR liegt vor, wenn die Verteilung von Z von den fehlenden Komponenten ζ_{fehl} abhängt. Damit ist z.B. gemeint, dass fehlende Werte einer Variable mit den Ausprägungen der Variable selbst in Zusammenhang stehen. Es gilt:

$$q(Z|\zeta, \phi) = q(Z|\zeta_{fehl}, \phi), \text{ für alle } \zeta_{fehl}, \phi, \quad (2.3)$$

(vgl. LITTLE und RUBIN, 2002, S. 12).

Im Rahmen dieser Arbeit ist gemäß der allgemeinen Formulierung von MAR und NMAR die folgende Abgrenzung von Bedeutung. Es wird davon ausgegangen, dass fehlende Werte nur innerhalb der Beobachtungsvariable y auftreten, jedoch nicht innerhalb der Hilfsvariablen. Daher wird von MAR gesprochen, wenn der Ausfall von den vollständig beobachteten Hilfsvariablen x_k für $k = 1, \dots, K$ abhängig ist, jedoch nicht vom Untersuchungsmerkmal y , welche fehlende Werte enthält. Ist der Ausfall von der nicht vollständig beobachteten, interessierenden Variablen y abhängig wird dieser Ausfallmechanismus als NMAR bezeichnet.

Die Fälle NMAR und MAR können des Weiteren noch unterschieden werden, wie sich die fehlenden Werte in Abhängigkeit der Ausprägung der Variablen verhalten. So könnten fehlende Werte zum Beispiel eher bei sehr hohen oder sehr niedrigen Werten des interessierenden Merkmals auftreten (vgl. LONGFORD, 2005, S. 31 f.).

An dieser Stelle ist zudem noch auf eine Einteilung aufmerksam zu machen, wie sie insbesondere in den Arbeiten von RUBIN (1987) oder LITTLE und RUBIN (2002) vorgenommen wird. Grundsätzlich kann ein Ausfallmechanismus als ignorierbar oder nicht ignorierbar beschrieben werden. Von einem ignorierbaren Mechanismus wird gesprochen, wenn die Eigenschaften einer gegebenen Inferenzstatistik nicht von dem zugrunde liegenden Mechanismus beeinflusst wird. Ansonsten wird er als nicht ignorierbar bezeichnet (vgl. BERGER et al., 2004, S. 3).

Weiterhin können fehlende Werte durch die Unterscheidung *Item-* und *Unit-Nonresponse* charakterisiert werden. Im Falle von Item-Nonresponse antworten Einheiten bezüglich bestimmter Fragen nicht. Fehlende Werte liegen für eine bestimmte Einheit nur hinsichtlich einzelner Variablen vor. Im Gegensatz hierzu fällt die Einheit bei Unit-Nonresponse komplett aus (vgl. LITTLE und RUBIN, 2002, S. 5 f.). Im Rahmen dieser Arbeit wurde zuvor die Annahme getroffen, dass in den Hilfsvariablen keine fehlenden Werte auftauchen. Daher wird hier von Item-Nonresponse ausgegangen. Bestimmte Imputationsmethoden können jedoch auch im Falle von Unit-Nonresponse angewendet werden.

2.2 Imputation fehlender Werte

Eine Möglichkeit zur Kompensation von fehlenden Werten, welche im Zentrum dieser Arbeit steht, ist die Imputation fehlender Werte. Die Imputation bezeichnet den *Prozess der Zuweisung von (konstruierten oder existierenden) Werten zu fehlenden Werten einer Variable unter einem möglichen Rückgriff beobachtbarer Hilfsvariablen zur Generierung eines vollständigen Datensatzes* (vgl. HAZIZA, 2009, S. 215, KIM und RAO, 2009, S. 917, CHAUVET et al., 2011, S. 459).

Es sollen Werte für fehlende Werte gefunden werden, die den tatsächlichen Ausprägungen möglichst nahe kommen. SÄRNDAL und LUNDSTRÖM (2005, S. 44) sprechen hier auch von *Proxy-Values*. Zusätzlich bezeichnen sie die imputierten Werte als eine Art Schätzung für die tatsächlichen Werte.

Mit der Imputation werden verschiedene Ziele verfolgt. Nach KIM und RAO (2009) soll die Verzerrung, welche mit fehlenden Werten einhergeht, der sogenannte NR-Bias, reduziert werden. Das damit verbundene Ziel ist die Realisation von unverzerrten Punktschätzungen. Dabei soll die tatsächliche Verteilung der zu imputierenden Variable möglichst gut wiedergegeben werden. Dies betrifft auch die Zusammenhänge zwischen verschiedenen Variablen (vgl. hierzu auch HAZIZA, 2009, S. 216). Nach MÜNNICH et al. (2015) ist es für die Amtliche Statistik wichtig, realitätsnahe und plausible Imputationswerte zu erhalten. Damit sollen z.B. Ausprägungen wie berufstätige Kinder unter 15 vermieden werden (vgl. MÜNNICH et al., 2015, S. 285). Um die bisher genannten Ziele zu realisieren, ist es in der Praxis von entscheidender Bedeutung ein gutes Imputationsmodell mit geeigneten Hilfsvariablen aufzustellen.

Nach KIM und SHAO (2014, S. 917) ist es zudem das Ziel der Imputation, einen vollständigen Datensatz zu erhalten, damit die Punktschätzungen konsistent sind für verschiedene Nutzer. Außerdem kann die Arbeit in der Praxis deutlich erleichtert werden, da Standardmethoden häufig auf vollständigen Datensätzen basieren.

Weitere Ziele können insbesondere in Zusammenhang mit der Varianz eines Schätzers stehen und sind daher gerade im Rahmen dieser Arbeit von Bedeutung. So sollte die Punktschätzung möglichst effizient sein (vgl. SÄRNDAL et al., 1992, S. 83 in Verbindung mit HAZIZA, 2009, S. 215). In den folgenden Abschnitten wird noch zu sehen sein, dass bestimmte Imputationsmethoden einen varianzerhöhenden Effekt mit sich bringen. Außerdem, und dies ist der zentrale Punkt dieser Arbeit, soll im Falle von fehlenden Werten und damit einhergehender Imputation eine unverzerrte Varianzschätzung resultieren. Dies ist dadurch möglich, dass alle Komponenten der Varianz berücksichtigt werden (vgl. HAZIZA, 2010, MASHREGHI et al., 2014, HAZIZA, 2009, S. 215 und SÄRNDAL, 1992).

In der Praxis ist es weiterhin häufig wichtig, dass die Imputationmethode nicht zu komplex und nicht zu zeitintensiv ist. Ziel ist es, die Imputationen in einer gewissen Zeit und vertretbarem Aufwand durchzuführen. Dies ist auch in Zusammenhang mit der Varianzschätzung von Bedeutung. Aufwendige und komplexe Imputationen können auch zu sehr zeitintensiven Varianzschätzungen führen (vgl. hierzu auch MÜNNICH et al., 2015, S. 273 ff., sowie SÄRNDAL, 1992, S. 243).

Im Rahmen dieser Arbeit wird im Zusammenhang mit der Imputation auf die folgende Nomenklatur zurückgegriffen: grundsätzlich wird der imputierte Wert für einen fehlenden Wert $i \in M$ durch \tilde{y}_i beschrieben. M stellt die Menge der Einheiten mit fehlendem Wert dar. Diese Findung möglichst geeigneter Ausprägungen geschieht auf Basis der Einheiten, die bezüglich des Untersuchungsmerkmals geantwortet haben und für welche gilt $i \in R$. Eventuell erfolgt auch ein Rückgriff auf bestimmte Hilfsvariablen, bei welchen in dieser Arbeit angenommen wird, dass sie vollständig beobachtet worden sind. Ob

zusätzlich Hilfsinformationen berücksichtigt werden, hängt vom Imputationsverfahren ab. Die Durchführung von Imputationen erfolgt häufig für fehlende Werte einer bestimmten Stichprobe S (mit $S = R \dot{\cup} M$), welche nach einem bestimmten Stichprobendesign aus der Grundgesamtheit gezogen wurde.¹

2.2.1 Klassifikation von Imputationsmethoden

Zunächst können Imputationsverfahren dadurch unterschieden werden, ob ein oder mehrere Werte zur Berechnung des imputierten Wertes herangezogen werden. Dadurch ergibt sich die Unterscheidung in *Single-* und *Multiple-Imputation*. Es ergeben sich verschiedene Vor- und Nachteile für die beiden Vorgehensweisen.

RUBIN (1986) kritisierte an Single-Imputationsverfahren, dass die Verwendung eines einzigen Wertes als imputierten Wert die Unsicherheit, welcher Wert verwendet werden soll, nicht abdecken kann. Den einen richtigen Wert gibt es nach seiner Ansicht nicht, da der Wert sonst auch nicht fehlen würde. SÄRNDAL (1992) hebt in diesem Zusammenhang hervor, dass Multiple-Imputation die Idee der Erzeugung von Variabilität durch die Imputation kommuniziert. Die Methode unterscheidet die Variabilität zwischen und innerhalb der erzeugten Datensätze. Dies wird durch die Varianzschätzmethoden, basierend auf Multiple-Imputation, gezeigt. Für SÄRNDAL (1992) bedeutet dies aber nicht zwangsläufig, dass Single-Imputationsmethoden diese zusätzliche Variabilität nicht richtig erfassen können.

Als Nachteile für Multiple-Imputationsverfahren wird hingegen oft angeführt, dass diese nicht praktikabel oder schwer durchzuführen sind und daher häufig nicht akzeptiert werden (vgl. SÄRNDAL, 1992, S. 243). Häufig wird die Plausibilität der erhaltenen Werte in Frage gestellt² und sie kommen daher in der Amtlichen Statistik noch nicht in dem Maße zum Einsatz wie die Single-Imputationsmethoden.

Im Rahmen dieser Arbeit wird daher der Fokus insbesondere auf die Verfahren der Single-Imputation gelegt. Es wird zudem auch eine Imputationsmethode untersucht, welche als *Fractional-Imputation* bezeichnet wird und derzeit sehr stark in der Wissenschaft gegenüber den klassischen Ansätzen der Multiple-Imputation diskutiert wird. Dieses Verfahren kann streng genommen nicht mehr zu den Verfahren der Single-Imputation gezählt werden, da hier mehrere Werte zur Berechnung des Imputationswertes für einen fehlenden Wert berücksichtigt werden. Es wird daher in dieser Arbeit generell von Imputation und nicht von Single-Imputation gesprochen.

Eine weitere Unterscheidung zwischen den Imputationsmethoden wird oftmals über das zugrunde liegende Imputationsmodell vorgenommen. Will ein Analyst eine Imputationsregel festlegen, realisiert er dies über die Wahl eines Modells. Die Imputationsregel gibt die Beziehung zwischen den Variablen einer Untersuchung, also insbesondere des Untersuchungsmerkmals und den Hilfsvariablen, wieder. Und hier ergeben sich zwei Möglichkeiten. So kann die Imputationsmethode das Modell explizit oder implizit verwenden. Ersteres ist z.B. bei Verhältnis-, Mittelwert- oder Regressionsimputation der Fall. Hier wird das Modell von *außen* mitgegeben, also schon vor der Durchführung des Verfahrens. Andere Methoden wie *Hot-Deck-Random* oder *Nearest-Neighbour-Imputation* berücksichtigen

¹ Im Rahmen dieser Arbeit werden die Imputationen auf Basis der Stichprobe bzw. im Rahmen der Varianzschätzung über Resampling-Verfahren in der Substichprobe vorgenommen. Bei Letzterem erfolgen jedoch andere Bezeichnungen, vgl. Abschnitt 4.1.2. Fehlende Werte können in der Grundgesamtheit und in der Stichprobe entstehen, vgl. hierzu Abschnitt 3.2

² Es ist jedoch auch bei Multiple-Imputationsverfahren möglich Plausibilitätsregeln einzuhalten. Ein gutes Beispiel sind die in MÜNNICH et al. (2015) im Rahmen der Simulationsstudie berücksichtigten Multiple-Imputationsverfahren.

Deterministische Imputationsverfahren	Zufällige Imputationsverfahren
Mittelwertimputation	Hot-Deck-Rand.-Imp. (HDRI)
Verhältnisismputation	Stoch. Regressionsimputation
Regressionsimputation	Fractional-Imputation
Nearest-Neighbour-Imputation	
Predictive-Mean-Matching	

Tabelle 2.1: Einteilung der Imputationsverfahren

ein Modell nur implizit (vgl. SÄRNDAL, 1992, S. 242 ff.). Im Rahmen dieser Arbeit werden Imputationsmethoden verwendet, welche sowohl auf expliziten und impliziten Modellen basieren. Dabei ist es ebenfalls möglich, dass eine Methode beide Modelltypen berücksichtigen kann. Ein Beispiel hierfür ist *Predictive-Mean-Matching*, welches in Abschnitt 2.2.7 näher erläutert wird.

Die wohl wichtigste Unterscheidung der verschiedenen Imputationsverfahren, insbesondere im Hinblick auf die Varianzschätzung, ist die Einteilung nach deterministischen und zufälligen Imputationsverfahren. Bei deterministischen Imputationsverfahren werden meist alle beobachtbaren Einheiten der interessierenden Variable sowie geeignete Hilfsvariablen zur Bestimmung des imputierten Wertes verwendet. Im Falle zufälliger Imputationsverfahren werden Zufallsstichproben von beobachtbaren Werten oder Residuen gezogen, um fehlende Werte zu imputieren (vgl. SHAO und SITTER, 1996, S. 1278). Eine bessere Abgrenzung wird in SÄRNDAL und LUNDSTRÖM (2005) gegeben. So handelt es sich bei deterministischen Imputationsmethoden um Verfahren, bei welchen der imputierte Wert bei einer gegebenen Stichprobe s und gegebenen beobachtbaren und fehlenden Werten auch bei Wiederholung der Imputation immer zum gleichen Imputationswert führt. Bei zufälligen Imputationsverfahren hingegen kann sich der Imputationswert bei Wiederholung der Imputation ändern, da die Auswahl dieses Wertes zufällig erfolgt (vgl. SÄRNDAL und LUNDSTRÖM, 2005, S. 153). Nach SHAO und STEEL (1999, S. 264) kann ein deterministisches Imputationsverfahren durch einfache Addition einer zufälligen Größe u_i in ein zufälliges Imputationsverfahren überführt werden. Tabelle 2.1 gibt die Einteilung der Imputationsverfahren an, wie sie in dieser Arbeit verwendet werden. Für Mittelwertimputation, Verhältnisismputation, Predictive-Mean-Matching und Nearest-Neighbour-Imputation werden in der Arbeit ausschließlich die deterministische Variante berücksichtigt. Der Grund liegt darin, dass die Eigenschaften zufälliger Imputationsverfahren bezüglich der Varianzschätzung anhand Methoden wie der Hot-Deck-Random-Imputation, der stochastischen Regressionsimputation sowie Fractional-Imputation gezeigt werden. Bei den vier genannten Methoden, bei welchen ausschließlich die deterministische Variante berücksichtigt wird, stehen hingegen die originären Eigenschaften dieser Methoden für die Varianzschätzung im Vordergrund. Aber auch diese Verfahren könnten grundsätzlich um einen zufälligen Term erweitert werden.

Grundsätzlich sprechen für die Verwendung von deterministischen oder zufälligen Imputationsverfahren verschiedene Vor- und Nachteile, bzw. Imputationsverfahren werden häufig bezüglich bestimmter Eigenschaften miteinander verglichen. Solche Vergleiche beziehen z.B. darauf,

- inwieweit die Verteilung des Untersuchungsmerkmals y gestört wird (vgl. hierzu z.B. LANDERMAN et al., 1997, S. 7, CHAUVET et al., 2011, S. 459, HAZIZA, 2009, S. 216 oder LITTLE und RUBIN, 2002, S. 61 f.),

- inwieweit die Variabilität der Schätzstatistik zunimmt (vgl. hierzu z.B. LANDERMAN et al., 1997, S. 7 oder CHAUVET et al., 2011, S. 460),
- ob tatsächlich beobachtete oder konstruierte Werte verwendet werden (vgl. z.B. CHEN und SHAO, 2000, S. 114 oder MÜNNICH et al., 2015, S. 272 f.),
- inwieweit Hilfsvariablen im Rahmen des Imputationsmodells bei der Berechnung der Imputationswerte eingehen (vgl. hierzu z.B. LANDERMAN et al., 1997, S. 6, CHEN und SHAO, 2000, S. 114 oder MÜNNICH et al., 2015, S. 272 f.),
- welche Folgen Verletzungen der Modellannahmen mit sich bringen (vgl. hierzu CHEN und SHAO, 2000, S. 114),
- wie komplex bzw. rechenintensiv die Imputationsmethode ist (vgl. hierzu z.B. die Ausführungen von MÜNNICH et al. (2015, S. 273 ff.) zur Komplexität von Nearest-Neighbour-Imputation und der damit zusammenhängenden Varianzschätzung).

Bei einigen deterministischen Imputationsverfahren stehen die Nachteile einiger Methoden gerade im Zusammenhang mit dem Punkt der Störung der Verteilung des Untersuchungsmerkmals y bzw. der daraus resultierenden geringeren Variabilität der erhaltenen Imputationswerte (vgl. hierzu z.B. LANDERMAN et al., 1997, S. 7, CHAUVET et al., 2011, S. 459 oder LITTLE und RUBIN, 2002, S. 61 f.). Dieser Punkt wird im Rahmen der einzelnen Imputationsmethoden näher erläutert. Außerdem können, wie im Falle von Verhältnis- oder Regressionsimputation, keine design- oder modellkonsistente Stichprobenquantile resultieren (vgl. SHAO et al., 1998, S. 821). Wie LANDERMAN et al. (1997, S. 7) hingegen für zufällige Imputationsverfahren hervorheben, wird durch Hinzufügen einer zufälligen Komponente die Varianz, welcher durch die Imputation entsteht, durch einen weiteren Unsicherheitsfaktor erhöht und die Präzision der Imputation reduziert. Weitere Vor- und Nachteile beziehen sich auf die einzelnen Verfahren, welche im Folgenden näher erläutert werden.³

2.2.2 Mittelwertimputation

Dieses Verfahren stellt eine einfache Form der Imputation dar. Für den fehlenden Wert j einer Einheit wird hier der Mittelwert der r aufgezeichneten Einheiten bezüglich des Untersuchungsmerkmals imputiert:

$$\tilde{y}_j = \bar{y}_R = \frac{1}{r} \sum_{i \in R} y_i. \quad (2.4)$$

Dadurch ergibt sich als Schätzer im Falle der einfachen Zufallsstichprobe für den Mittelwert der Grundgesamtheit:

$$\hat{y}_{MW} = 1/n \cdot \left(\sum_{i \in R} y_i + \sum_{i \in M} \tilde{y}_i \right) = 1/n \cdot (r \cdot \bar{y}_R + m \cdot \bar{y}_R) = \bar{y}_R, \quad (2.5)$$

wobei $m = n - r$ die Anzahl der fehlenden Einheiten darstellt.

³ Ein Vergleich verschiedener Imputationsmethoden wird z.B. in LITTLE und RUBIN (2002), LANDERMAN et al. (1997), HAZIZA (2009) oder LONGFORD (2005) gegeben.

Dieser Einfachheit der Durchführung stehen gravierende Nachteile gegenüber. Durch die Mittelwertimputation wird die empirische Verteilung des zu ermittelten Untersuchungsmerkmals y gestört. Durch die Reduktion der Imputation auf den Mittelwert geht die Variabilität der Untersuchungsvariable verloren. Statistische Größen wie die Varianz, die Schiefe oder auch Perzentile, welche nicht als einfache Linearkombination der Daten berechnet werden, können durch die Anwendung von Standardmethoden nicht konsistent geschätzt werden (vgl. LITTLE und RUBIN, 2002, S. 61 f., SÄRNDAL und LUNDSTRÖM, 2005, S. 168 f.).

Weiterhin bemängeln LANDERMAN et al. (1997), dass Hilfsinformationen von anderen Variablen des Datensatzes zur Verbesserung der Genauigkeit der Imputation bei ausschließlichen Rückgriff auf die Mittelwerte nicht berücksichtigt werden. Zudem werden auch Korrelationen zu anderen interessierenden Variablen nicht einbezogen, was zu einer Abschwächung dieser Korrelation führt und auch Verzerrungen mit sich bringt (vgl. LANDERMAN et al., 1997, S. 6).

Eine Verbesserung dieser Methode ist die konditionierte Mittelwertimputation, durch welche Informationen von Hilfsvariablen einbezogen werden können. Hier erfolgt die Einteilung der Einheiten auf Basis der Hilfsvariablen x_k für $k = 1, \dots, K$ in Imputationsklassen oder Gruppen g . Für die fehlenden Werte einer Klasse wird der Mittelwert der aufgezeichneten Werte dieser Klasse genommen (vgl. LITTLE und RUBIN, 2002, S. 61 f.). Die beschriebenen Probleme liegen weiterhin in abgeschwächter Form innerhalb der einzelnen Imputationsklassen vor.

2.2.3 Verhältnisimputation

Bei dieser Imputationsmethode⁴ wird der imputierte Wert auf Basis eines vorliegenden Hilfsmerkmals x berechnet. Hierzu muss der Koeffizient $\hat{\beta}$ berechnet werden, welcher sich aus dem Verhältnis der Summen der Beobachtungs- und Hilfsvariablen für die vollständig aufgezeichneten Einheiten ergibt: $\hat{\beta} = \frac{\sum_{i \in R} y_i}{\sum_{i \in R} x_i}$. Aufgrund dieses Vorgehens wird vorausgesetzt, dass die Hilfsvariable metrisch ist. Der imputierte Wert für einen fehlenden Wert $j \in M$ errechnet sich anschließend aus:

$$\tilde{y}_j = \hat{\beta} \cdot x_j. \quad (2.6)$$

Der Mittelwertschätzer \hat{y}_{Rat} kann im Falle der Verhältnisimputation und der einfachen Zufallsstichprobe hergeleitet werden gemäß:

$$\begin{aligned} \hat{y}_{Rat} &= \frac{1}{n} \cdot \left(\sum_{i \in R} y_i + \sum_{i \in M} \tilde{y}_i \right) \\ &= \frac{1}{n} \cdot \left(\sum_{i \in R} y_i + \hat{\beta} \cdot \sum_{i \in M} x_i \right) \\ &= \frac{1}{n} \cdot \left(\sum_{i \in R} y_i + \frac{\sum_{i \in R} y_i}{\sum_{i \in R} x_i} \cdot \sum_{i \in M} x_i \right) \\ &= \frac{1}{n} \cdot \left(\sum_{i \in R} y_i + \frac{\bar{y}_R}{\bar{x}_R} \cdot \sum_{i \in M} x_i \right) \\ &= \bar{y}_R \cdot \frac{1}{n} \cdot \left(r + (n - r) \cdot \bar{x}_M / \bar{x}_R \right) \end{aligned}$$

⁴ Verhältnisimputation wird im Rahmen dieser Arbeit abgekürzt durch Rat von der englischen Übersetzung *Ratioimputation*.

$$\begin{aligned}
 &= \bar{y}_R / \bar{x}_R \cdot \frac{1}{n} \cdot (r \cdot \bar{x}_R + (n - r) \cdot \bar{x}_M) \\
 &= \frac{\bar{y}_R}{\bar{x}_R} \cdot \bar{x}_S \\
 &= \hat{\beta} \cdot \bar{x}_S
 \end{aligned} \tag{2.7}$$

mit $\bar{x}_S = (1/n \cdot \sum_{i=1}^n x_i)$, $\bar{y}_R = (1/r \cdot \sum_{i \in R} y_i)$ und $\bar{x}_R = (1/r \cdot \sum_{i \in R} x_i)$. Damit stellt der Mittelwertschätzer bei Anwendung von Verhältnisimputation unter den gegebenen Bedingungen nichts anderes dar als der Verhältnisschätzer (vgl. hierzu auch LEE et al., 1994, S. 232).

Entsprechend gilt für den Totalwertschätzer $\hat{\tau}_{Rat}$:

$$\hat{\tau}_{Rat} = N \cdot \frac{\bar{y}_R}{\bar{x}_R} \cdot \bar{x}_S.$$

Beide Statistiken können im Falle von Verhältnisimputation relativ einfach und schnell auf Basis der vollständig beobachtbaren Hilfsvariable x und den antwortenden Einheiten für die Untersuchungsvariable y berechnet werden. Die Formel zeigt jedoch auch, dass die interessierende Statistik keine stochastischen Elemente enthält und damit bei Wiederholung der Schätzung auf Basis der gleichen Stichprobe zu gleichen Ergebnissen führt. Mit der Verhältnisimputation liegt damit wiederum ein deterministisches Imputationsverfahren vor, bei welchem die empirische Verteilung gestört und die tatsächliche Varianz des Untersuchungsmerkmals unterschätzt wird. Diese Unterschätzung ist jedoch nicht so stark wie bei der Mittelwertimputation. Bei letzterer wird für die fehlenden Werte ausschließlich der Mittelwert genommen. Bei der Verhältnisimputation unterscheiden sich die imputierten Werte \tilde{y}_j in Abhängigkeit der Hilfsvariable. So entsteht mehr Variation unter den imputierten Werten. Ein weiterer Nachteil der Verhältnisimputation liegt in der Begrenztheit ihres Anwendungsgebietes bei Durchführung in der beschriebenen Form. Zunächst ist sie auf eine Hilfsvariable beschränkt. Für die Berücksichtigung mehrerer Hilfsvariablen müssten die einzelnen Variablen mit Hilfe von Gewichtungen kombiniert werden, was jedoch eher unüblich ist. Außerdem erfordert die Bildung von Verhältnissen metrische Variablen, wodurch ihre Anwendung bei den, für die Praxis wichtigen kategorialen Variablen, nicht möglich ist.

2.2.4 Regressionsimputation

Zur Bestimmung eines imputierten Wertes wird hier auf das bekannte Regressionsmodell zurück gegriffen. Auch diese Imputationsmethode geht vom Vorhandensein geeigneter Hilfsvariablen x_k für $k = 1, \dots, K$ aus, für die vollständige Beobachtungen aller Befragten vorliegen. Für die interessierende Variable beobachtbaren Einheiten des Antwortsets $i \in R$ wird eine Regression durchgeführt mit y als abhängige Variable und den x_k als unabhängige Variablen. Die Regressionskoeffizienten α und β_k , $k = 1, \dots, K$ werden in diesem Zusammenhang bei Annahme eines linearen Zusammenhanges z.B. mittels der Methode der kleinsten Quadrate oder der Maximum-Likelihood-Methode geschätzt. Um die m fehlenden Werte von y zu schätzen, werden deren Beobachtungen bezüglich der einzelnen Hilfsvariablen anschließend in die Regressionsgleichung eingesetzt:

$$\tilde{y}_j = \hat{\alpha} + \sum_{k=1}^K \hat{\beta}_k \cdot x_{kj}, \text{ für alle } j = 1, \dots, m, \tag{2.8}$$

wobei $\hat{\alpha}$ das geschätzte Absolutglied und $\hat{\beta}_k$ die geschätzten Steigungskoeffizienten der Regression darstellen, welche auf Basis der vollständigen Fälle ermittelt wurden. Die Regression kann stetige oder kategoriale Variablen, Interaktionen oder weniger restriktive parametrische Formen wie *Splines* umfassen, um die Vorhersage zu verbessern (vgl. LITTLE und RUBIN, 2002, S. 62). Dadurch wird ihr Einsatzbereich gegenüber der Verhältnisimputation deutlich ausgeweitet.

Im Vergleich zur Mittelwertimputation in ihrer unkonditionierten Form sind hier weniger restriktive Annahmen bezüglich des Ausfallmechanismus gegeben. Bei der Regressionsimputation reicht es aus, dass der Mechanismus zumindest als MAR bezeichnet werden kann. Der Ausfallmechanismus sollte im Rubin'schen Sinne ignorierbar sein. Zudem wird im Vergleich zur Mittelwertimputation die Verteilung der interessierenden Variable y weniger stark gestört. Der Grund hierfür liegt darin, dass die imputierten Werte sich nicht mehr auf den Mittelwert konzentrieren und sich durch die Verwendung der geschätzten Werten aus dem Regressionsmodell \hat{y}_i eine gewisse Variation resultiert. Die Varianz wird weniger stark unterschätzt, da die Summe der quadratischen Abweichungen der \hat{y}_i um den Gesamtmittelwert in den Varianzschätzungen enthalten sind. Dennoch ist weiterhin eine gewisse Störung der Verteilung des Untersuchungsmerkmals y gegeben, da die imputierten Werte auf die \hat{y}_i begrenzt sind. Des Weiteren liegt eine Unterschätzung der Varianz vor, solange das Bestimmtheitsmaß der Regression kleiner als 1 ist (vgl. LANDERMAN et al., 1997, S. 7 und die darin aufgeführte Literatur). In Abschnitt 2.2.9 wird gezeigt, dass die Unterschätzung der Varianz der tatsächlichen Werte von y und damit auch die Störung ihrer Verteilung gerade mit der Korrelation von y und x zusammenhängt. Davon ist auch abhängig, ob in diesem Punkt gegenüber der Verhältnisimputation Vorteile erzielt werden. Auch bei Verwendung von Regressionsimputation im Falle der Anwendung der Methode der kleinsten Quadrate können Mittelwerte und Totalwerte einfach geschätzt werden. Zur besseren Übersichtlichkeit wird dies anhand einer Hilfsvariablen und der einfachen Zufallsstichprobe gezeigt (vgl. hierzu auch KIM, 2001 sowie KIM und SHAO, 2014, welche ebenfalls den Mittelwertschätzer basierend auf Regressionsimputation als Regressionsschätzer interpretieren):

$$\begin{aligned}
 \hat{y}_{Reg} &= \frac{1}{n} \cdot \left(\sum_{i \in R} y_i + \sum_{i \in M} \tilde{y}_i \right) \\
 &= \frac{1}{n} \cdot (r \cdot \bar{y}_R + m \cdot \hat{\alpha} + m \cdot \hat{\beta} \cdot \bar{x}_M) \\
 &= \frac{1}{n} \cdot (r \cdot \bar{y}_R + m \cdot (\bar{y}_R - \hat{\beta} \cdot \bar{x}_R) + m \cdot \hat{\beta} \cdot \bar{x}_M) \\
 &= \frac{1}{n} \cdot (n \cdot \bar{y}_R - m \cdot \hat{\beta} \cdot \bar{x}_R + m \cdot \hat{\beta} \cdot \bar{x}_M) \\
 &= \frac{1}{n} \cdot (n \cdot \bar{y}_R - m \cdot \hat{\beta} \cdot \bar{x}_R + \hat{\beta} \cdot (\sum_{i \in S} x_i - r \cdot \bar{x}_R)) \\
 &= \frac{1}{n} \cdot (n \cdot \bar{y}_R + \hat{\beta} \cdot \sum_{i \in S} x_i - m \cdot \hat{\beta} \cdot \bar{x}_R - r \cdot \hat{\beta} \cdot \bar{x}_R) \\
 &= \frac{1}{n} \cdot (n \cdot \bar{y}_R + \hat{\beta} \cdot \sum_{i \in S} x_i - n \cdot \hat{\beta} \cdot \bar{x}_R) \\
 &= (\bar{y}_R + \hat{\beta} \cdot \bar{x}_S - \hat{\beta} \cdot \bar{x}_R) \\
 &= \hat{\alpha} + \hat{\beta} \cdot \bar{x}_S.
 \end{aligned} \tag{2.9}$$

Entsprechend gilt für den Totalwertschätzer $\hat{\tau}_{Reg}$:

$$\hat{\tau}_{Reg} = N \cdot (\hat{\alpha} + \hat{\beta} \cdot \bar{x}_S).$$

Im Falle einer Hilfsvariable unterscheiden sich die Verhältnisimputation und die Regressionsimputation beim Mittelwert bzw. beim Totalwertschätzer nur hinsichtlich der Bestimmung der Parameter $\hat{\alpha}$ und $\hat{\beta}$. Jedoch ist es bei der Regressionsimputation im Gegensatz zur Verhältnisimputation einfacher möglich, mehrere Hilfsvariablen durch die Aufnahme in das Regressionsmodell zu berücksichtigen. In diesem Fall gilt für den Schätzer in (2.9):

$$\hat{y}_{Reg} = \hat{\alpha} + \hat{\beta} \cdot \bar{x}_S$$

wobei $\hat{\beta}$ und \bar{x}_S Vektoren mit den entsprechenden Regressionskoeffizienten bzw. Mittelwerten der Hilfsvariablen darstellen.

Um den Nachteilen einer deterministischen Regression zu begegnen, ist es zudem möglich, diese stochastisch zu gestalten. Dies geschieht dadurch, dass dem imputierten Wert aus (2.8) eine zufällige Größe u_i im Sinne einer Störterms hinzugefügt wird (vgl. z.B. LITTLE und RUBIN, 2002, S. 65):

$$\hat{y}_i = \hat{\alpha} + \sum_{k=1}^K \hat{\beta}_k \cdot x_{ki} + u_i, \text{ für alle } i = 1, \dots, m. \quad (2.10)$$

Diese Störgröße u_i kann z.B. aus den Residuen der antwortenden Einheiten der zugrunde liegenden Regression zufällig gezogen werden (vgl. KIM, 2001, S. 76). Oder die Größe stammt aus einer Normalverteilung mit einem Erwartungswert von null und der Varianz σ_u^2 , welche der Varianz der Störgröße der Regression von y auf x_k entspricht (vgl. LITTLE und RUBIN, 2002, S. 65). Weiterhin ist es auch denkbar, diese zufälligen Störgrößen auf Basis eines bestimmten Modells auszuwählen (vgl. DAVID et al., 1986, S. 34).

Durch die Berücksichtigung einer zufälligen Größe sind die imputierten Werte nicht mehr auf \hat{y}_i begrenzt. Da sowohl die Varianz zwischen und innerhalb der Ebenen von \hat{y}_i bei der Imputation von y_i berücksichtigt wird, ist es möglich, neben dem Mittelwert auch die Varianz der imputierten Variable zu schätzen (vgl. LANDERMAN et al., 1997, S. 7). Wie bei jedem zufälligen Imputationsverfahren wird aber, wie schon zuvor erläutert, die Varianz des Schätzers durch die Hinzunahme einer zufälligen Größe erhöht.

Sowohl bei der Regressionsimputation wie auch der stochastischen Regressionsimputation gilt es zu beachten, dass nicht nur reale, sondern unter Umständen auch nicht gewünschte Imputationswerte realisiert werden können. Ein Beispiel wäre, dass Einkommen imputiert werden und sich aufgrund des Regressionsmodells oder der Hinzunahme von Residuen negative Einkommen ergeben, welche in der Erhebung nicht vorgesehen sind. Eine Korrektur solcher Imputationswerte z.B. in Richtung Nulleinkommen kann mit Verzerrungen einhergehen. Solche Sachverhalte sollten über das Imputationsmodell berücksichtigt werden.

2.2.5 Hot-Deck-Random-Imputation

Im Kern besteht dieses Verfahren darin, fehlende Wert dahingehend zu kompensieren, dass aus einem bestimmten Pool von sogenannten Spendern, Imputationswerte nach dem Modell mit oder dem Modell ohne Zurücklegen zufällig gezogen werden. Die Spender

sind im allgemeinen Fall die Einheiten, für welche bezüglich des Untersuchungsmerkmals Beobachtungen vorliegen (vgl. hierzu LONGFORD, 2005, S. 43 f.).

Dieser Pool an Spender kann theoretisch alle Einheiten mit beobachtbarem Wert der Stichprobe umfassen oder über bestimmte Hilfsvariablen eingegrenzt werden, so dass gewährleistet ist, dass nur Spender verwendet werden, welche der Einheit mit fehlendem Wert sehr ähnlich sind. Im allgemeinen Fall gilt für den Mittelwert:

$$\bar{y}_{HDRI} = \{r \cdot \bar{y}_R + m \cdot \bar{y}_M^*\} / n \quad (2.11)$$

wobei \bar{y}_R den Mittelwert der Antwortenden bezeichnet und \bar{y}_M^* berechnet wird durch

$$\bar{y}_M^* = \sum_{i \in R} \frac{d_i \cdot y_i}{m}. \quad (2.12)$$

d_i gibt in diesem Zusammenhang an, wie oft der Wert einer Auskunftsperson des Pools gezogen wurde (vgl. LITTLE und RUBIN, 2002, S. 66 f.).

Die Hot-Deck-Random-Imputation (HDRI) ist ein klassisches zufälliges Imputationsverfahren. Dadurch besitzt auch dieses Verfahren den großen Vorteil, z.B. gegenüber der Mittelwertimputation, dass die Verteilung des Untersuchungsmerkmals y nicht durch die Imputation gestört wird (vgl. LITTLE und RUBIN, 2002, S. 66 f.). Aber auch hier nimmt die Variabilität des Schätzers zu. Ein weiterer Vorteil der Methode ergibt sich daraus, dass nur tatsächlich realisierte Werte verwendet werden. Es erfolgt eine Zufallsauswahl aus den auftretenden Werten der antwortenden Einheiten.

In der einfachen Hot-Deck-Random-Imputation werden keine Hilfsvariablen berücksichtigt, sondern die Spender werden aus dem Pool der antwortenden Einheiten gezogen. Dies kann in Abhängigkeit des Antwortmechanismus zu starken Verzerrungen und inkonsistenten Schätzern führen. Daher ist es gerade bei diesen Methoden sinnvoll, Imputationsklassen über die Hilfsvariablen zu bilden, welche mit den fehlenden Werten in Verbindung stehen. Spender für einen fehlenden Wert können dann nur noch aus der gleichen Imputationsklasse stammen (vgl. LITTLE und RUBIN, 2002, S. 68 f.).

In der Literatur finden sich zahlreiche Versionen von Hot-Deck-Random-Imputation. So ist es z.B. möglich, den Spendern Gewichte auf Basis von Distanzfunktionen zuzuordnen und damit deren Auswahlwahrscheinlichkeiten unterschiedlich zu wählen (vgl. LONGFORD, 2005, S. 43 ff.). Denkbar wäre es, wie z.B. in RAO und SHAO (1992) oder SHAO et al. (1998, S. 821) beschrieben, die Auswahlwahrscheinlichkeiten der Spender proportional zu den Designgewichten zu bilden.⁵ Im Falle konstanter Antwortwahrscheinlichkeiten innerhalb der Imputationsklassen resultieren bei Anwendung der gewichteten Hot-Deck-Random-Imputation designkonsistente Schätzungen von Totalwerten und Stichprobenquantilen (vgl. SHAO et al., 1998, S. 821).

Weiterhin können sich die Verfahren hierbei unterscheiden, ob die Spender mit oder ohne Zurücklegen gezogen werden. Für beide Varianten sprechen verschiedene Vor- und Nachteile. Für das Ziehen mit Zurücklegen spricht der Vorteil, dass immer ausreichend Spender vorhanden sind, auch bei großen Ausfallraten. Dies gilt insbesondere dann, wenn Imputationsklassen gebildet worden sind. Durch das Ziehen mit Zurücklegen wird gewährleistet, dass ein optimaler Spender gewählt wird und nicht aus einer anderen Imputationsklasse ausgewählt werden muss, falls in einer Imputationsklasse nicht genügend Spender zur

⁵ Wie in MASHREGHI et al., 2014 beschrieben, wird die Auswahlwahrscheinlichkeit eines Spenders in diesem Fall durch dessen Designgewicht geteilt durch die Summe der Designgewichte aller Spender berechnet.

Verfügung stehen. Für das Ziehen ohne Zurücklegen spricht hingegen, dass bestimmte Spender nicht zu oft gezogen und als Imputationswert verwendet werden. Dies gilt gerade für Extremwerte, welche unter Umständen zu oft ausgewählt werden. Eine grundlegende Aussage für die eine oder andere Variante ist daher schwer zu treffen und hängt häufig vom konkreten Untersuchungsgegenstand ab (vgl. hierzu auch die Ausführungen zur Problematik bezüglich der Bestimmung von Spendern mit und ohne Zurücklegen im Falle von Nearest-Neighbour-Imputation in MÜNNICH et al., 2015, S. 290 f.).

2.2.6 Nearest-Neighbour-Imputation

Dieses Verfahren kann als konkreter Spezialfall der Hot-Deck-Random-Imputation aufgefasst werden (vgl. zu den Ausführungen dieses Abschnittes auch MÜNNICH et al., 2015, S. 272 ff.). Hierbei wird für den fehlenden Wert der Wert einer antwortenden Person genommen, die dem Verweigerer bezüglich bestimmter Hilfsvariablen am ähnlichsten ist. Bei metrischen Hilfsvariablen kann auf ein metrisches Maß zurückgegriffen werden (vgl. LITTLE und RUBIN, 2002, S. 69 ff.). Die einfachste Möglichkeit, einen solchen nächsten Nachbarn zu bestimmen, ist, basierend auf den Hilfsvariablen x_k für $k = 1, \dots, K$, den einfachen absoluten Abstand zu berechnen. Dabei wird bei einer Hilfsvariable für einen fehlenden Wert y_j der beobachtete Wert y_i imputiert, wenn für diesen gilt:

$$|x_i - x_j| = \min_{l \in R} |x_l - x_j|. \quad (2.13)$$

Diese Vorgehensweise kann auch bei Vorhandensein mehrerer Hilfsvariablen angewendet werden. Es müssen dann entsprechende Gewichtungen und Standardisierungen vorgenommen werden (vgl. CHEN und SHAO, 2000, S. 114).

Aus der Berechnung von (2.13) resultiert die sogenannten Distanzmatrix. Die Zeilen bilden die beobachteten Einheiten und die Spalten die Einheiten mit fehlendem Wert. Die Länge der Spalten bzw. Zeilen wird durch die entsprechenden Anzahlen definiert.

Statt der absoluten Distanz bietet es sich ebenfalls an, andere mathematische Maße zur Bestimmung des nächsten Nachbarn zu nehmen. Denkbar wäre es z.B., die Mahalanobis Distanz oder die euklidische Distanz zu verwenden (vgl. LITTLE und RUBIN, 2002, S. 69). Handelt es sich bei den Hilfsvariablen um kategoriale Variablen, müssen komplexere Maße benutzt werden, welche solche Variablentypen berücksichtigen. Beispiele hierfür sind die Gower-Distanz und der Simple-Matching-Coefficient.⁶ Zudem kann es gerade bei kategorialen Hilfsvariablen häufig vorkommen, dass mehrere nächste Nachbarn mit dem gleichen Wert einer Hilfsvariable existieren. CHEN und SHAO (2000) schlagen in solchen Fällen vor, eine Zufallsauswahl aus diesen Einheiten vorzunehmen. Dies ist dem allgemeinen Vorgehen der Hot-Deck-Random-Imputation bei Verwendung von Imputationsklassen sehr ähnlich. Als Vorteile der Nearest-Neighbour-Imputation gegenüber Verfahren wie der Regressionsimputation, der Mittelwertimputation oder der Verhältnisimputation wird häufig angeführt, dass hier keine konstruierten Werte, sondern tatsächlich realisierte Werte für die fehlenden Werte verwendet werden. Außerdem verzichtet dieses Verfahren explizit auf bestimmte Modellannahmen. Dies macht das Verfahren robust gegenüber Verletzungen der Modellannahmen, wie sie bei modellbasierten Verfahren wie z.B. Regressionsimputation und Verhältnisimputation auftreten können. Zudem wird im Artikel von CHEN und

⁶ Zum Beispiel verwendet die Nearest-Neighbour-Imputation des R-Paketes VIM die Gower-Distanz und die des `scrime`-Paketes den Simple-Matching-Coefficient (vgl. TEMPL et al., 2016, SCHWENDER und FRITSCH, 2013, sowie SCHWENDER und ICKSTADT, 2008).

SHAO (2000) zeigt, dass die Verteilung der interessierenden Variablen y im Vergleich zur Regressionsimputation, der Mittelwertimputation oder der Verhältnisimputation nicht gestört wird, wodurch die beschriebenen Verzerrungen nicht auftreten können. Dennoch ist es gegenüber Verfahren wie z.B. der einfachen Hot-Deck-Random-Imputation effizient, da auf Hilfsvariablen zurückgegriffen wird und es sich um ein nicht komplett zufälliges Verfahren handelt (vgl. CHEN und SHAO, 2000, S. 114 sowie MÜNNICH et al., 2015, S. 272 f.). Als problematisch bei der Nearest-Neighbour-Imputation kann sich die Berechnung der Distanzen ergeben. Dies ist z.B. der Fall bei großen Stichprobengrößen bzw. Imputationsklassen. Hier ergeben sich große Distanzmatrizen, da viele Distanzen zwischen den beobachteten und fehlenden Werten zu berechnen sind. Dadurch kann die Anwendung der Methode deutlich erschwert oder sogar unmöglich werden. Ein gutes Beispiel einer Untersuchung bei welcher aufwendige Distanzmatrizen zu berechnen sind, ist der Deutsche Zensus (vgl. MÜNNICH et al., 2015, S. 273 ff.).

Ein weiteres Problem der Nearest-Neighbour-Imputation wird in LONGFORD (2005) angesprochen. Demnach ist die Qualität der Imputation für die Einheiten mit fehlendem Wert sehr unterschiedlich. Manche dieser Einheiten besitzen sehr ähnliche nächste Nachbarn. Andere jedoch sind sehr stark isoliert und der nächste Nachbar besitzt nur geringe Ähnlichkeiten. Dadurch kann es vorkommen, dass manche Spender sehr oft für die Imputation herangezogen werden und manche überhaupt nicht (vgl. LONGFORD, 2005, S. 43). Dies hat natürlich auch Auswirkungen auf die Variabilität der imputierten Variable. Wenn jedoch eine möglichst gleichmäßige Inanspruchnahme der Spender stattfindet, kann hingegen eine gute Reproduktion der Variabilität erfolgen. In diesem Zusammenhang ist es auch bei diesem Verfahren möglich, die Inanspruchnahme von beobachteten Werten zu begrenzen (vgl. hierzu auch MÜNNICH et al., 2015, S. 290).

Der Mittelwertschätzer beziehungsweise Totalwertschätzer im Falle von Nearest-Neighbour-Imputation kann nach CHEN und SHAO (2000) wie im Falle der Hot-Deck-Random-Imputation sehr einfach beschrieben werden durch:

$$\bar{y}_{NN} = \frac{1}{n} \sum_{i \in R} (1 + d_i) \cdot y_i,$$

wobei d_i angibt wie oft ein Element $i \in R$ als nächster Nachbar verwendet wurde. Für den Totalwertschätzer gilt:

$$\bar{y}_{NN} = \frac{N}{n} \sum_{i \in R} (1 + d_i) \cdot y_i.$$

2.2.7 Predictive-Mean-Matching

Eine weitere Imputationsmethode ist Predictive-Mean-Matching (vgl. im Folgenden LITTLE, 1986, LANDERMAN et al., 1997). Diese stellt eine Kombination aus Regressionsimputation und Nearest-Neighbour-Imputation dar. Im Folgenden wird von einer metrischen interessierenden Variable y ausgegangen. Wie bei der Regressionsimputation wird auch hier ein Regressionsmodell nach (2.8) aufgestellt, welches auf Basis der beobachteten Werte geschätzt wird. Hierbei werden die imputierten Werte aber nicht nach dem Regressionsmodell berechnet. Vielmehr werden über das Regressionsmodell die Werte der interessierenden Variablen für die beobachteten und fehlenden Werte geschätzt, um über diese anschließend den nächsten Nachbarn gemäß (2.13) zu finden:

$$|\hat{y}_i - \hat{y}_j| = \min_{l \in R} |\hat{y}_l - \hat{y}_j|. \quad (2.14)$$

Für einen fehlenden Wert wird damit jener beobachtbare Wert verwendet, dessen prognostizierter Wert \hat{y}_i nach dem Regressionsmodell (2.8) dem entsprechenden prognostizierten Wert des Nichtantwortenden nach einer bestimmten Distanzfunktion am nächsten kommt. In (2.14) wird wieder der absolute Abstand verwendet.

Diese Vorgehensweise hat den Vorteil, dass es bei einem gut funktionierendem Regressionsmodell sehr robust ist. Insbesondere können dadurch neben mehreren Hilfsvariablen auch ihre Zusammenhänge und Interaktionen bei der Distanzberechnung berücksichtigt werden. Dies kann ein Vorteil gegenüber der einfachen Nearest-Neighbour-Imputation sein, bei der eine Berücksichtigung dieser Komponenten unter Umständen deutlich aufwendiger ist. Wird von einem metrischen interessierendem Merkmal und damit auch prognostizierten Werten dieses Skalenniveaus ausgegangen, können auch die einfachen metrischen Distanzfunktionen verwendet werden. Weiterhin ergeben sich die zuvor beschriebenen Vorteile der Nearest-Neighbour-Imputation, wie z.B. die gute Reproduktion der Verteilung und die Inanspruchnahme tatsächlich realisierter Werte.

Jedoch bringt dieses Verfahren ebenfalls die Nachteile großer Distanzmatrizen mit sich und Modellverletzungen können durch die expliziten Annahmen größere Folgen haben als bei der einfachen Nearest-Neighbour-Imputation. Insbesondere die Auswahl des optimalen Nachbarn kann dadurch beeinträchtigt werden.

Eine Zuordnung von Predictive-Mean-Matching entweder zu deterministischen oder zufälligen Imputationsverfahren ist abhängig davon, welches Regressionsmodell dem Verfahren zugrunde liegt. So wäre es ebenso denkbar, dem Regressionmodell einen Störterm hinzuzufügen, wie es bei stochastischer Regressionsimputation in Gleichung (2.10) der Fall ist. In diesem Fall ist der nächste Nachbar auch von dieser zufälligen Störgröße abhängig und kann sich bei sonst gleichen Bedingungen bei Wiederholung der Imputation ändern. Wird hingegen, wie zuvor erläutert, das für eine einfache Regressionimputation verwendete Modell (2.8) benutzt, ist der nächste Nachbar eindeutig festgelegt. Auch bei Wiederholung der Imputation bei den gegebenen Daten wird dieser wiederholt resultieren. In diesem Fall ist Predictive-Mean-Matching den deterministischen Imputationsverfahren zuzuordnen. In dieser Arbeit wird die letztere Variante berücksichtigt.

2.2.8 Fractional-Imputation

Diese Methode ist insbesondere durch die Arbeiten von KALTON und KISH (1984), FAY (1996), KIM und FULLER (2004), FULLER und KIM (2005) und KIM und SHAO (2014) bekannt geworden. In der Literatur tritt diese Variante häufig in Kombination mit Hot-Deck-Random-Imputation auf. Im Gegensatz zu der einfachen Hot-Deck-Random-Imputation werden hierbei für jeden fehlenden Wert j , mit $j \in M$, anstelle von einem Spender eine bestimmte Anzahl von D Spendern im Modell mit Zurücklegen gezogen. Jedem dieser D Spender wird ein bestimmte *Fraction* \tilde{w}_{ij} des ursprünglichen Gewichtes der Einheit mit fehlendem Wert zugeordnet, welches im einfachsten Fall $\tilde{w}_{ij} = 1/D$ beträgt. Nach KIM und FULLER (2004) müssen die Fractions der D Spender, mit $i \in R$, so konstruiert werden, dass gilt:

$$\sum_{i=1}^D \tilde{w}_{ij} = 1. \quad (2.15)$$

Der imputierte Wert \tilde{y}_j , mit $j \in M$, wird anschließend berechnet durch:

$$\tilde{y}_j = \sum_{i=1}^D \tilde{w}_{ij} \cdot y_i. \quad (2.16)$$

Für $\tilde{w}_{ij} = 1/D$ resultiert:

$$\tilde{y}_j = \frac{1}{D} \cdot \sum_{i=1}^D y_i \quad (2.17)$$

und der imputierte Wert ist das einfache arithmetische Mittel über die D Spender. Bei der hier beschriebenen Variante von Fractional-Imputation handelt es sich um eine sehr einfache Form. Komplexere Formen beschäftigen sich insbesondere mit der optimalen Bestimmung der Fractions \tilde{w}_{ij} . Beispiele finden sich in der am Anfang dieses Abschnittes gegebenen Literatur. Es werden häufig auch Imputationsklassen gebildet, und die Bestimmung der Fractions in jeder Klasse separat durchgeführt (vgl. z.B. FULLER und KIM, 2005). Im Rahmen dieser Arbeit wird ausschließlich die einfache Form verwendet.

Durch die enge Verbindung zur Hot-Deck-Random-Imputation und da durch die Wiederholung der Imputation auf die gleiche Stichprobe verschiedene Imputationswerte resultieren können, kann diese einfache Form der Fractional-Imputation den zufälligen Imputationsverfahren zugeordnet werden. Jedoch geht durch die Durchschnittsbildung der D Imputationswerte wie bei der Mittelwertimputation Variance verloren. Die Varianz der Imputationswerte kann dadurch deutlich geringer sein als die tatsächliche Varianz. In Abschnitt 2.2.9 wird sich zeigen, dass auf diesem Wege auch Eigenschaften der Mittelwertimputation auftreten können.

Nach FULLER und KIM (2005, S. 140 f.) bzw. KIM und FULLER (2004, S. 561 f.) kann ein linearer Schätzer auf Basis von Fractional-Imputation relativ einfach angegeben werden durch

$$\hat{\theta}_{\text{frac}} = \sum_{i \in R} \left(\sum_{j \in S} w_j \cdot \tilde{w}_{ij} \right) \cdot y_i = \sum_{i \in R} \rho_i \cdot y_i, \quad (2.18)$$

wobei $\rho_i = \sum_{j \in S} w_j \cdot \tilde{w}_{ij}$. Wenn $i = j$ eine antwortende Einheit ist, nimmt \tilde{w}_{ij} eins an. Für $i \neq j$ mit $i \in R$ und $j \in R$ ist $\tilde{w}_{ij} = 0$. Ansonsten gelten für $i \neq j$ mit $i \in R$ und $j \in M$ die Bedingungen der Bestimmung der Fractions nach (2.15).

2.2.9 Varianzen der imputierten Werte der verschiedenen Imputationsmethoden

Die Abbildungen 2.1 und 2.2 zeigen die Verteilungen der imputierten Werte, welche nach den beschriebenen Methoden der vorherigen Abschnitten imputiert wurden. Die Bandbreite der erhaltenen Imputationswerte wird in Tabelle 2.2 gegeben. Die Ergebnisse sind für eine konkrete Stichprobe angegeben und können daher für detailliertere Analysen nicht uneingeschränkt verwendet werden. Sie sollen aber die zuvor getätigten Aussagen an konkreten Daten demonstrieren. In diesem Zusammenhang spielen insbesondere die originären Eigenschaften der Methoden auf die Verteilung der imputierten Werte eine Rolle. Daher soll an dieser Stelle von anderen Fragestellungen wie z.B. die Bildung von Imputationsklassen abstrahiert werden. Solche Themen werden im späteren Verlauf der Arbeit aufgegriffen.

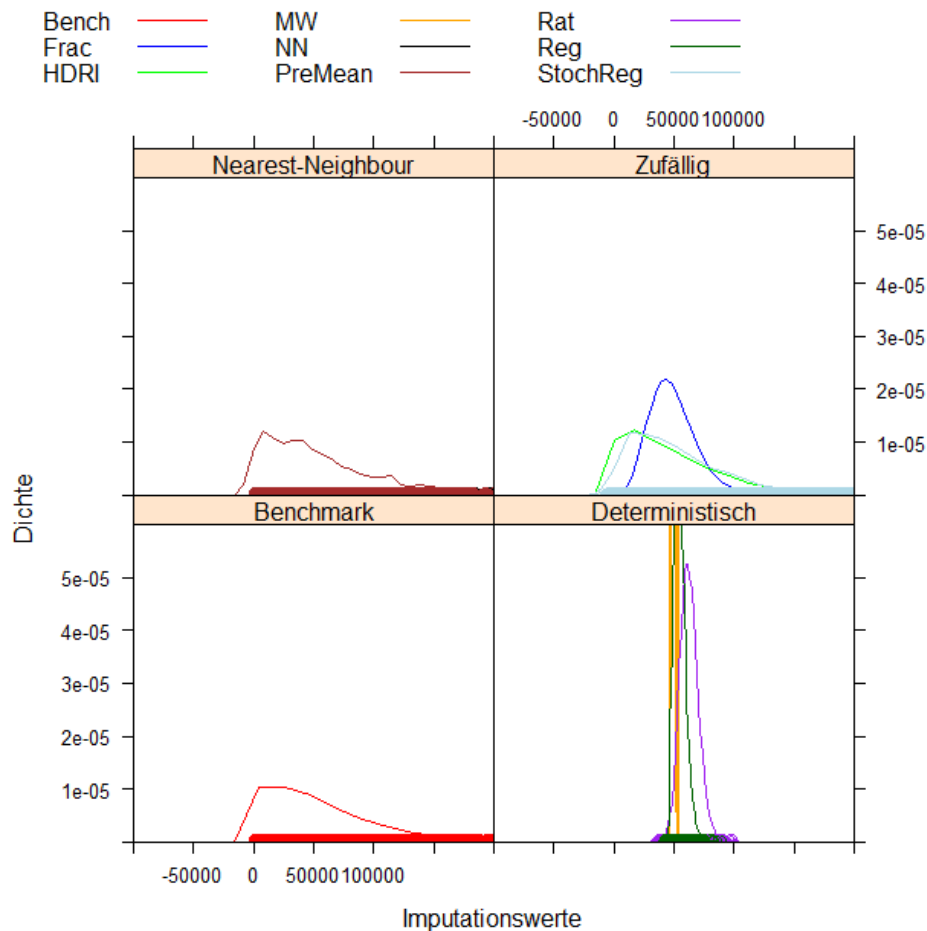


Abbildung 2.1: Verteilungen der imputierten Werte der einzelnen Methoden in der Grundgesamtheit 1 bei geringer Korrelation

In diesem Zusammenhang werden zwei verschiedene Grundgesamtheiten benutzt.⁷ Die interessierende Variable ist die aufsummierte Einkommensvariable aus dem Datensatz AMELIA, beschrieben in 5.1.1. Die Grundgesamtheiten unterscheiden sich hinsichtlich der Hilfsvariable x . Diese ist metrisch und wird über ein Modell erzeugt (siehe Modell (5.1) in Abschnitt 5.1.1), in Abhängigkeit der Korrelation von x und y .⁸ So ist in der ersten Grundgesamtheit in Abbildung 2.1 die Korrelation zwischen den beiden Variablen gering (bei 11 %). In Abbildung 2.2 ist sie hingegen sehr groß (bei 70 %). Grundgesamtheit 1 beschreibt ein klassisches MAR Szenario. Bei Grundgesamtheit 2 hingegen ist durch den hohen Zusammenhang zum Untersuchungsmerkmal eine Mischform bezüglich MAR und NMAR gegeben, da fehlende Werte auch eine gewisse Abhängigkeit zum Untersuchungsmerkmal besitzen.

Die fehlenden Werte sind für alle Grundgesamtheiten synthetisch in Abhängigkeit der Hilfsvariable nach einem bestimmten Modell (vgl. Abschnitt 5.1.4) erzeugt worden. Die tatsächlichen Ausprägungen dieser Einheiten sind daher bekannt. Ihre Verteilung dient als Benchmark und ist in allen Abbildungen in der linken unteren Box durch die rot gefärbte Verteilung angegeben. Die Verteilungen der imputierten Werte, welche durch die

⁷ Vgl. hierzu die Parameterkonstellationen in der Datensatzbeschreibung in Abschnitt A.2.2.1 in Anhang A.

⁸ Die beiden Hilfsvariablen verwenden zwar das gleiche Modell wie die in Abschnitt 5.1.1 beschriebene metrische Hilfsvariable, entsprechen dieser jedoch nicht.

verschieden Methoden realisiert wurden, sind in den drei übrigen Quadranten dargestellt. Die Imputationsmethoden wurden dahingehend zusammen abgebildet, ob sie den deterministischen, den zufälligen oder den Nearest-Neighbour-Verfahren⁹ zuzurechnen sind. Bei letzteren handelt es sich in dieser Arbeit zwar auch um deterministische Verfahren, aus Gründen der Übersichtlichkeit werden sie jedoch separat abgebildet.

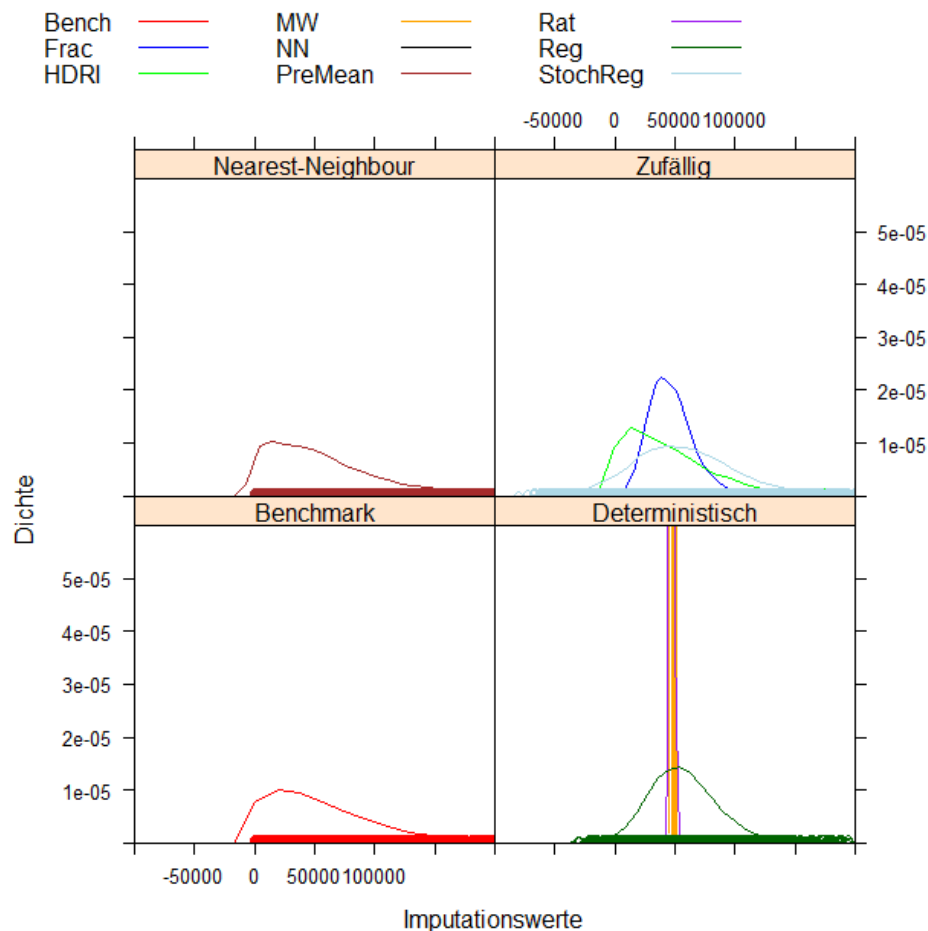


Abbildung 2.2: Verteilungen der imputierten Werte der einzelnen Methoden in der Grundgesamtheit 2 bei hoher Korrelation

Bei den deterministischen Verfahren ist zu sehen, dass durch Mittelwert- und Verhältnisimputation (*MW* bzw. *Rat*) die Verteilung der imputierten interessierenden Variable sehr stark auf wenige Werte begrenzt ist. Dies gilt sowohl für Grundgesamtheit 1 wie auch für Grundgesamtheit 2. Ein Vergleich mit dem Benchmark zeigt, dass die Varianz der Untersuchungsvariablen durch diese Imputationen stark unterschätzt wird. Dies war schon in den theoretischen Ausführungen erwartet worden.

Bei Regressionsimputation (*Reg*) hängt das Ergebnis sehr stark von der Grundgesamtheit ab. Im Fall einer geringen Korrelation des Untersuchungsmerkmals und damit einem schwächeren Zusammenhang des Regressionsmodells gleicht die Verteilung derjenigen von Mittelwert - bzw. Verhältnisimputation. Damit ist die Verteilung sehr stark begrenzt und die Varianz der tatsächlichen Werte stark unterschätzt. Anders sieht es bei hoher Korrelationen in Grundgesamtheit 2 aus. Hier wird die Verteilung weniger stark gestört und ist wesentlich flacher. Die Varianz wird nicht mehr so stark unterschätzt. In Abschnitt 2.2.4

⁹ Hierunter wird neben der Nearest-Neighbour-Imputation auch Predictive-Mean-Matching gefasst.

wurde schon kurz darauf eingegangen, dass das Ausmaß der Unterschätzung der Varianz der imputierten Werte bei Regressionsimputation sehr stark vom Bestimmtheitsmaß der Regression abhängt. Unter Vorbehalt weiterer Einflussfaktoren kann dies auch für diese beiden Grundgesamtheiten festgestellt werden. Eine deutlich größere Annäherung an die wahre Verteilung wird mittels stochastischer Regressionsimputation (*StochReg*) erreicht. Durch die zufälligen Residuen wird hier mehr Variabilität in den Imputationswerten erzeugt, wie an beiden Grundgesamtheiten zu sehen ist. Problematisch ist bei beiden Methoden das Auftreten negativer Werte bei höherer Korrelation von y und x , wie in Tabelle 2.2 zu sehen ist. Die zu imputierende Variable bildet das Einkommen und der geringste Wert dieser Variable in der vollständigen Grundgesamtheit ist das Nulleinkommen. Die durch das Regressionsmodell auftretenden negativen Werte sind daher Ausprägungen, welche in der Grundgesamtheit nicht realisiert werden können. Begrenzungen des Imputationsverfahrens auf bestimmte Bereiche zu realisierender Imputationswerte können jedoch weitere Probleme mit sich bringen. Werden z.B. negative Werte einfach als Nulleinkommen behandelt, können Überschätzungen der interessierenden Statistik die Folge sein. Der Grund liegt darin, dass das verwendete Regressionsmodell so aufgebaut ist, dass die beobachteten Werte des Untersuchungsmerkmals y im Mittel abgedeckt werden. Werden die negativen Werte nicht mehr berücksichtigt, können Verzerrungen einhergehen. Dieser Sachverhalt kann daher gegen die Anwendung der deterministischen oder stochastischen Regressionsimputation bei Vorliegen von Variablen mit beschränktem Wertebereich sprechen. Da im Rahmen dieser Arbeit insbesondere die Varianzschätzung im Vordergrund steht, werden die negativen Werte auch als solche behandelt. Die Varianzschätzung soll dadurch nicht von bestimmten Verzerrungen des Punktschätzers beeinflusst werden.

Methode	Grundgesamtheit 1		Grundgesamtheit 2	
	Min.	Max.	Min.	Max.
Benchmark	0	1.018.000	0	864.000
MW	46.350	53.390	44.990	50.760
Rat	57.320	102.400	42.230	63.270
Reg	51.380	90.370	-34.930	386.900
NN	0	367.400	0	492.100
PreMean	0	367.400	0	492.100
StochReg	-12.490	623.800	-108.600	409.400
HDRI	0	750.700	0	644.800
Frac	2.107	188.700	2.291	182.500

Tabelle 2.2: Bandbreite der Imputationswerte

An der Verteilung der Imputationswerte basierend auf Fractional-Imputation (*Frac*) in den Abbildungen 2.1 und 2.2 wird die Mischform dieser Methode klar. Zwar zeigt die Verteilung, dass durchaus eine Variation der Imputationswerte gegeben ist. Jedoch wird diese durch die Durchschnittsbildung über die D Spender wieder stark begrenzt, so dass sich ihre Form wiederum jener für Mittelwertimputation annähert. Mit Hot-Deck-Random-Imputation (*HDRI*) wird die Verteilung der interessierenden Variable gut wiedergegeben und die tatsächliche Varianz kaum unterschätzt. Jedoch gilt es hier zu beachten, dass bei dieser Methode in ihrer ursprünglichen Form keine Hilfsvariablen berücksichtigt werden. Da der Ausfallmechanismus MAR ist, kann es durchaus sein, dass der Imputationswert konkreter Einheiten nicht nahe an der tatsächlichen Ausprägung liegt.

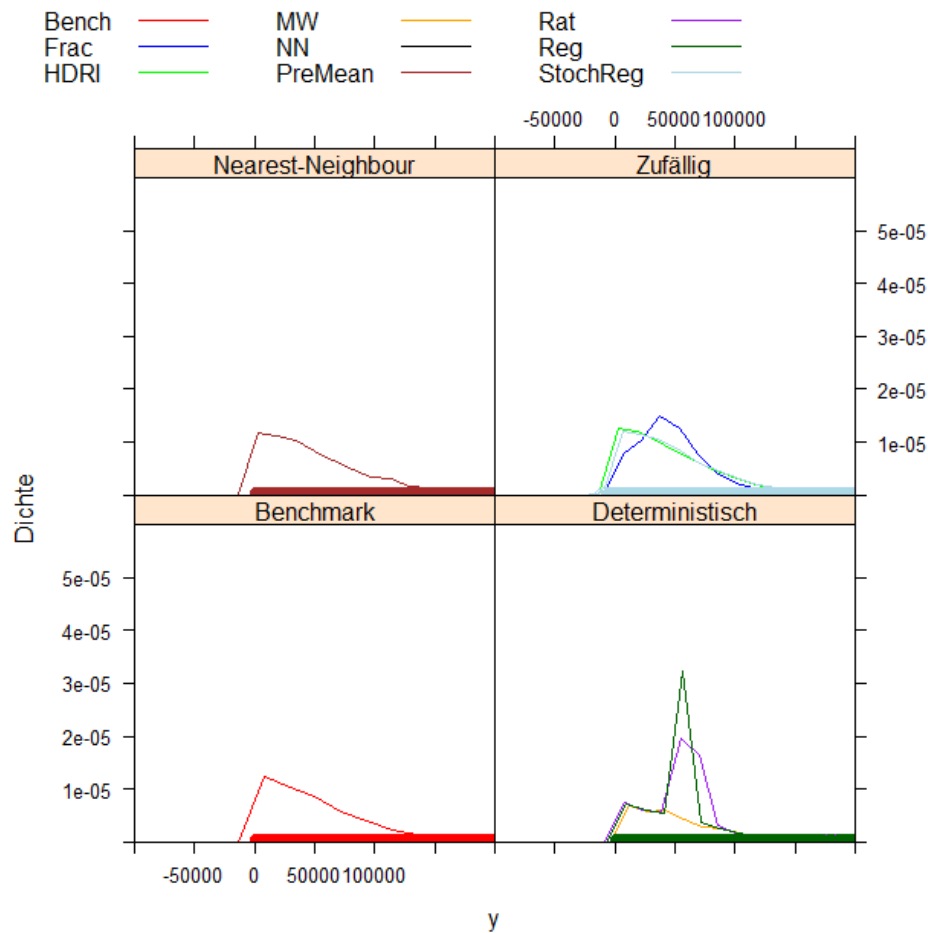


Abbildung 2.3: Verteilungen des Untersuchungsmerkmals y nach der Imputation mit den verschiedenen Methoden in der Grundgesamtheit 1 bei geringer Korrelation

Der Vergleich mit dem Benchmark in den Abbildungen 2.1 und 2.2 zeigt, dass auch Nearest-Neighbour-Imputation (NN) und Predictive-Mean-Matching ($PreMean$) die tatsächliche Verteilung der zu imputierenden Werte sehr gut wiedergeben und diese nicht stören. Die Imputationen dieser beiden Varianten sind in diesem Fall sogar identisch. Der Grund liegt darin, dass für alle Imputationen nur eine metrische Variable berücksichtigt worden ist. Die Hilfsvariable x im Falle der Nearest-Neighbour-Imputation stellt in diesem Fall eine einfache lineare Transformation über das Regressionsmodell beim Predictive-Mean-Matching dar. Unter sonst gleichen Bedingungen müssen die beiden Imputationsmethoden zu gleichen Ergebnissen führen.

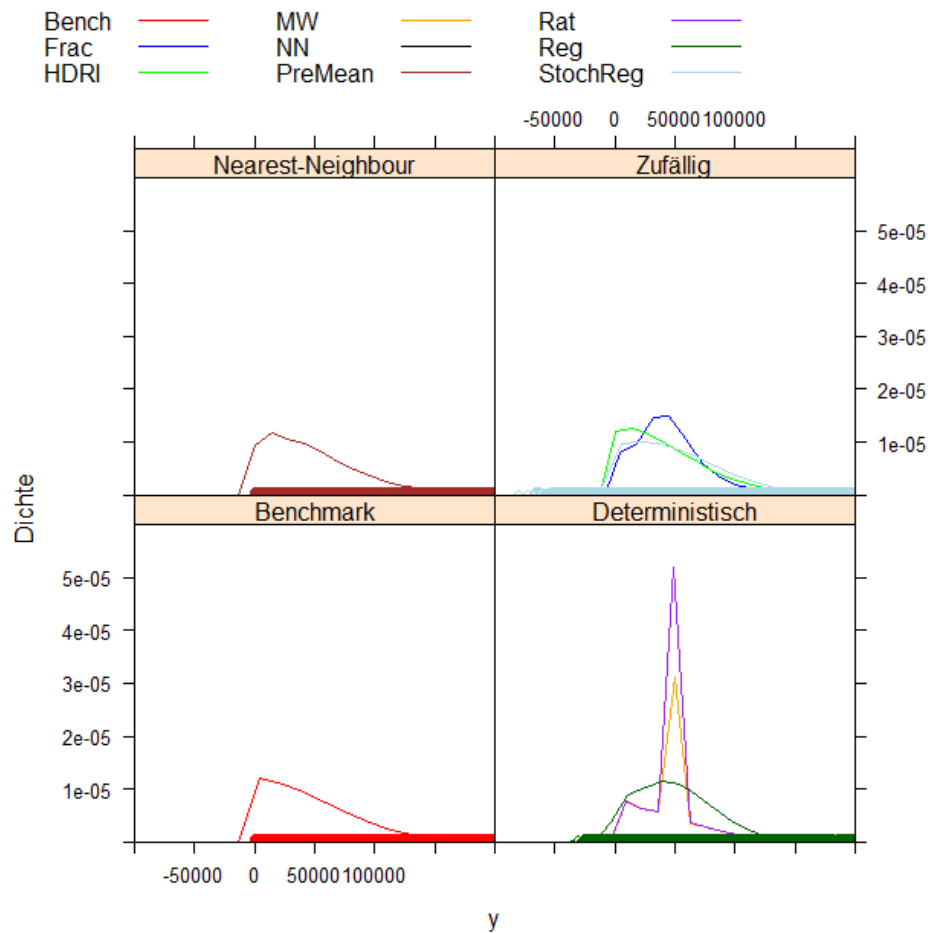


Abbildung 2.4: Verteilungen des Untersuchungsmerkmals y nach der Imputation mit den verschiedenen Methoden in der Grundgesamtheit 2 bei hoher Korrelation

Der Grund für die gute Wiedergabe der Verteilung der ursprünglichen Werte beider Imputationsmethoden liegt insbesondere darin, dass tatsächlich realisierte Werte genommen werden und in diesem Fall die Werte nicht auf einen bestimmten Bereich begrenzt werden wie im Falle der Verhältnis-, Mittelwert- oder Regressionsimputation. Es scheint, dass bestimmte Spender nicht zu oft als Imputationswert verwendet werden, sondern eine gleichmäßige Inanspruchnahme der beobachteten Werte gegeben ist. Durch die geeignete Wahl der Hilfsvariable und daraus resultierender nächster Nachbarn wird hier die Bandbreite der Werte der interessierenden Variable gut abgedeckt, mit Ausnahme mancher Ausreißer, wie in Tabelle 2.2 zu sehen ist. Zur Beurteilung der Imputationsmethoden ist es wichtig auch zu untersuchen, inwieweit ihre Imputationswerte die Verteilung des interessierenden Merkmals insgesamt beeinflussen. Daher werden in den Abbildungen 2.3 bzw. 2.4 die nach der Imputation resultierende Verteilung des Untersuchungsmerkmals dargestellt. Insbesondere bei den deterministischen Verfahren im unteren linken Quadranten ist zu erkennen, dass die Begrenzung der Imputationswerte auf wenige Werte auch die Verteilung des Untersuchungsmerkmals insgesamt stört.

2.2.10 Zusammenfassende Gegenüberstellung der Imputationsmethoden

In Tabelle 2.3 werden die relevanten Imputationsmethoden dieser Arbeit noch einmal bezüglich bestimmter Eigenschaften miteinander verglichen. Die Tabelle bezieht sich insbesondere auf die theoretischen Ausführungen der letzten Abschnitte. Ein X steht für Trifft (eher) zu und ein — für Trifft (eher) nicht zu. Die Spalte Hilfsvariable beschreibt, wie Hilfsvariablen bei der Imputationsmethode berücksichtigt werden. Zum Beispiel geschieht dies bei Nearest-Neighbour-Imputation über die Distanzberechnung und bei Hot-Deck-Random-Imputation über die Imputationsklassen (IKL).

Methode	Störung der Verteilung	Zus. Varianzkomponente	Realer Wert	Hilfsvariable	Komplexität / Rechenzeit
MW	X	—	—	IKL	gering
Rat	X	—	—	Modell	gering
Reg	X	—	—	Modell	gering
NN	—	—	X	Distanz	hoch
PreMean	—	—	X	Modell/ Distanz	hoch
StochReg	—	X	—	Modell	gering
HDRI	—	X	X	IKL	gering
Frac	X	X	—	IKL	eher gering

Tabelle 2.3: Gegenüberstellung der Imputationsmethoden

3 Varianzschätzung unter fehlenden Werten und Imputation

Das vorherige Kapitel beschäftigte sich schwerpunktmäßig mit den verschiedenen Antwortmechanismen und Imputationsmethoden. Varianzen wurden bis jetzt nur im Zusammenhang mit Imputationswerten betrachtet. Im Rahmen dieser Arbeit sind jedoch insbesondere die Varianzen bestimmter Schätzstatistiken unter fehlenden Werten und Imputation von Interesse. Dieser Sachverhalt steht im Fokus dieses Kapitels. Neben der Varianz und ihren Komponenten werden auch verschiedene direkte und modellassistierende Methoden zur Schätzung dieser Varianz angegeben werden. Diese werden hinsichtlich der in Abschnitt 2.2 beschriebenen Imputationsmethoden unterschieden. Zwei weitere, wichtige Bestimmungsfaktoren in diesem Zusammenhang sind der Entstehungsort fehlender Werte sowie der Response-Vektor z . Bevor sich der Thematik fehlender Werte und Imputation gewidmet wird, erfolgt die Angabe von Varianzen und ihrer Schätzer bei vollständigen Beobachtungen für die relevanten Stichprobendesigns dieser Arbeit.

Für die Erläuterungen dieses Kapitels wird auf lineare Statistiken wie den Totalwert τ oder den Mittelwert \bar{y} abgestellt. Diese werden geschätzt auf Basis einer Stichprobe S mittels des klassischen Horvitz-Thompsons-Schätzers (vgl. HORVITZ und THOMPSON, 1952). Unter Auslassung spezieller Indizes für die Designkomponenten und bei vollständigen Beobachtungen gilt:

$$\hat{\tau} = \sum_{i \in S} w_i \cdot y_i \quad (3.1)$$

$$\hat{\bar{y}} = \sum_{i \in S} \frac{w_i \cdot y_i}{\sum_{i \in S} w_i}, \quad (3.2)$$

wobei w_i die Designgewichte darstellen. Erfolgen die Schätzungen auf der imputierten Variable y_i^I mit

$$y_i^I := \begin{cases} y_i, & \text{wenn } i \in R, \\ \tilde{y}_i, & \text{wenn } i \in M, \end{cases} \quad (3.3)$$

statt auf den vollständigen Beobachtungen y_i , werden die Schätzer als $\hat{\tau}_I$ und $\hat{\bar{y}}_I$ bezeichnet. Verallgemeinert werden die beiden Schätzer durch die Verwendung von $\hat{\theta}$ und $\hat{\theta}_I$ als Schätzer für θ . Falls sich die Imputation auf eine spezielle Methode bezieht, wird statt I im Index die Abkürzung des Verfahrens verwendet.

Nichtlineare Statistiken wie Armutssmaße oder Kalibrierungsansätze wie der GREG-Schätzer werden nicht betrachtet. Hier sind unter Umständen spezielle Linearisierungen notwendig, welche in dieser Arbeit nicht behandelt werden.

3.1 Varianzschätzung für Totalwertschätzer im Rahmen verschiedener Stichprobendesigns bei vollständigen Beobachtungen

Der grundlegendste Fall eines Stichprobendesigns ist die einfache Zufallsstichprobe (*Simple-Random-Sampling*: SRS).¹ Hier wird aus der Grundgesamtheit der Größe N eine Stichprobe der Größe n gezogen. Im Folgenden wird vom Modell ohne Zurücklegen ausgegangen. Im Falle von vollständigen Beobachtungen kann die Varianz des Totalwertschätzers angegeben werden durch:

$$\begin{aligned} V(\hat{\tau}_{SRS}) &= N^2 \cdot \left(\frac{N-n}{N} \right) \cdot \frac{\sigma^2}{n}, \\ \text{mit } \sigma^2 &= \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y}_U)^2, \\ \text{und } \hat{\tau}_{SRS} &= \sum_{i=1}^n \frac{N}{n} \cdot y_i = \sum_{i=1}^n w_i \cdot y_i, \end{aligned} \quad (3.4)$$

wobei w_i definiert ist durch $w_i = N/n$ und $\left(\frac{N-n}{N} \right)$ die Endlichkeitskorrektur darstellt, welche im Falle des Ziehens ohne Zurücklegen relevant ist.²

Der erwartungstreue Standardvarianzschätzer für (3.4) ist gegeben durch:

$$\begin{aligned} \hat{V}(\hat{\tau}_{SRS}) &= N^2 \cdot \left(\frac{N-n}{N} \right) \cdot \frac{s^2}{n}, \\ \text{mit } s^2 &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned} \quad (3.5)$$

(vgl. LOHR, 1999, S. 29 ff.).

Ein etwas komplexeres Stichprobendesign ist die geschichtete Zufallsstichprobe (*Stratified-Random-Sampling*: StrRS). Hier wird die Grundgesamtheit in H Schichten eingeteilt, die aus insgesamt N_1, N_2, \dots, N_H Einheiten bestehen. Aus den einzelnen Schichten werden Stichproben der Größe n_1, n_2, \dots, n_h gezogen. Im Falle der geschichteten Zufallsstichprobe gilt für die Varianz:

$$\begin{aligned} V(\hat{\tau}_{StrRS}) &= \sum_{h=1}^H N_h^2 \cdot \left(\frac{N_h - n_h}{N_h} \right) \cdot \frac{\sigma_h^2}{n_h}, \\ \text{mit } \sigma_h^2 &= \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{y}_{U_h})^2, \\ \text{mit } \hat{\tau}_{StrRS} &= \sum_{h=1}^H \sum_{i=1}^{n_h} \frac{N_h}{n_h} \cdot y_{hi} = \sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} \cdot y_{hi}, \end{aligned} \quad (3.6)$$

¹ Vgl. zu diesem Abschnitt insbesondere LOHR, 1999, S. 29 ff., SÄRNDAL et al., 1992, S. 124 ff. aber auch die Ausführungen in BRUCH, 2010.

² Im Rahmen dieser Arbeit werden beide Schreibweisen $\sum_{i=1}^n y_i$ und $\sum_{i \in S} y_i$ als Summe über die Einheiten der Stichprobe äquivalent verwendet.

mit $w_{hi} = N_h/n_h$. Ein erwartungstreuer Varianzschätzer für (3.6) ist definiert durch:

$$\hat{V}(\hat{\tau}_{StrRS}) = \sum_{h=1}^H N_h^2 \cdot \left(\frac{N_h - n_h}{N_h} \right) \cdot \frac{s_h^2}{n_h}, \quad (3.7)$$

$$\text{mit } s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2,$$

(vgl. LOHR, 1999, S. 99 ff.).

Bei der klassischen zweistufigen Zufallsstichprobe ist die Grundgesamtheit U aufgeteilt in insgesamt L Primäreinheiten (*Primary-Sampling-Unit(s)*: PSU) $U_1, \dots, U_d, \dots, U_L$. Diese PSU setzen sich wiederum aus den Einheiten der zweiten Stufe (SSU, *Secondary-Sampling-Unit(s)*) der Größe $N_1, \dots, N_d, \dots, N_L$ zusammen. Bei diesem Design bilden die SSU auch die Einheiten der letzten Stufe (USU, *Ultimate-Sampling-Unit(s)*). Auf der ersten Stufe wird aus den L PSU eine einfache Zufallsauswahl von l PSU durchgeführt. Aus den so gewonnen Einheiten werden auf der zweiten Stufe n_d Einheiten aus den N_d SSU durch einfache Zufallsauswahl ausgewählt. Die Beobachtungswerte werden mit y_{di} bezeichnet (vgl. auch LOHR, 1999, S. 145 f., SÄRNDAL et al., 1992, S. 144).

Die Varianz des Totalwertschätzers $V(\hat{\tau}_{2St})$ im Fall der zweistufigen Zufallsstichprobe besteht aus zwei Komponenten: zum einen entsteht Variabilität zwischen den PSU und zum anderen ergibt sich Variabilität innerhalb der PSU mit Bezug auf die SSU. Die Varianz der zweistufigen Zufallsstichprobe resultiert damit aus der Varianz der ersten Stufe addiert um die Varianz der zweiten Stufe (ein ausführlicher Beweis zur Herleitung der wahren Varianz des Totalwertschätzers findet sich in LOHR, 1999, S. 204 ff.) und es gilt:

$$V(\hat{\tau}_{2St}) = L^2 \cdot \left(\frac{L-l}{L} \right) \cdot \frac{\sigma_\tau^2}{l} + \frac{L}{l} \sum_{d=1}^L \left(\frac{N_d - n_d}{N_d} \right) \cdot N_d^2 \cdot \frac{\sigma_d^2}{n_d}, \quad (3.8)$$

$$\text{mit } \sigma_\tau^2 = \frac{1}{L-1} \sum_{d=1}^L \left(\tau_d - \frac{\tau}{L} \right)^2,$$

$$\sigma_d^2 = \frac{1}{N_d - 1} \cdot \sum_{i=1}^{N_d} (y_{di} - \bar{y}_{Ud})^2,$$

$$\text{und } \hat{\tau}_{2St} = \frac{L}{l} \cdot \sum_{d=1}^l \sum_{i=1}^{n_d} \frac{N_d}{n_d} \cdot y_{di} = \sum_{d=1}^l \sum_{i=1}^{n_d} w_{di} \cdot y_{di},$$

mit $w_{di} = \frac{L}{l} \cdot \frac{N_d}{n_d}$. Dabei ist τ_d der Totalwert in der PSU d , geschätzt durch:

$$\hat{\tau}_d = \sum_{i=1}^{n_d} \frac{N_d}{n_d} \cdot y_{di}.$$

Der erste Term aus (3.8) ist die Varianz auf der ersten Stufe und beinhaltet σ_τ^2 die Varianz zwischen den Totalwerten der einzelnen PSU und σ_d^2 ist die Varianz zwischen den einzelnen Einheiten in der PSU d . Der zweite Term aus (3.8), die Varianz auf der zweiten Stufe, ist der Varianz im Fall der geschichteten Zufallsstichprobe gemäß Formel (3.6) sehr ähnlich. Es wird jedoch die Variabilität innerhalb der PSU und nicht der Schichten gemessen. Außerdem findet eine Hochrechnung der Varianz um den Quotienten L/l statt (vgl. LOHR, 1999, S. 147).

Als direkter Varianzschätzer kann folgender erwartungstreuer Schätzer verwendet werden:

$$\begin{aligned}\hat{V}(\hat{\tau}_{2St}) &= L^2 \cdot \left(\frac{L-l}{L}\right) \cdot \frac{s_\tau^2}{l} + \frac{L}{l} \sum_{d=1}^l \left(\frac{N_d - n_d}{N_d}\right) \cdot N_d^2 \cdot \frac{s_d^2}{n_d}, \\ \text{mit } s_\tau^2 &= \frac{1}{l-1} \sum_{d=1}^l \left(\hat{\tau}_d - \frac{\hat{\tau}_{2St}}{L}\right)^2, \\ \text{und } s_d^2 &= \frac{1}{n_d - 1} \cdot \sum_{i=1}^{n_d} (y_{di} - \bar{y}_d)^2,\end{aligned}\tag{3.9}$$

(vgl. LOHR, 1999, S. 147).

Häufig wird in der Literatur, insbesondere auch bei den später noch vorzustellenden Resampling-Methoden, bei der Varianzschätzung nur die erste Stufe berücksichtigt. Der Hintergrund liegt darin, dass mit den Varianzschätzern der einzelnen Stufen nicht die korrespondierende Varianz der Stufe geschätzt wird. Wie in SÄRNDAL et al. (1992) und LOHR (1999) beschrieben, wird mit dem Varianzschätzer der ersten Stufe deutlich mehr geschätzt als die Varianz der ersten Stufe. So gilt für dessen Erwartungswert:

$$\begin{aligned}\mathbb{E} \left[L^2 \cdot \left(\frac{L-l}{L}\right) \cdot \frac{s_\tau^2}{l} \right] &= L^2 \cdot \left(1 - \frac{l}{L}\right) \cdot \frac{\sigma_\tau^2}{l} \\ &\quad + \frac{L}{l} \left(1 - \frac{l}{L}\right) \sum_{d=1}^L \left(\frac{N_d - n_d}{N_d}\right) \cdot N_d^2 \cdot \frac{\sigma_d^2}{n_d},\end{aligned}\tag{3.10}$$

(vgl. LOHR, 1999, S. 209 f.).

Daher kann es ausreichen, nur die erste Stufe des Stichprobendesigns bei der Varianzschätzung zu berücksichtigen. Dies gilt nur unter ganz bestimmten Bedingungen. Nach SÄRNDAL et al. (1992), BRUCH (2010) und BRUCH et al. (2011) hängt dies vom Auswahlatz auf den verschiedenen Stufen sowie vom Varianzanteil der zweiten Stufe ab.

Der erste Bestimmungsfaktor ist der Auswahlatz auf der ersten Stufe. Der Vergleich von (3.10) mit der tatsächlichen Varianz beim zweistufigen Stichprobendesign in (3.8) zeigt, dass ein umso größerer Anteil der Gesamtvarianz abgedeckt wird, je kleiner der Auswahlatz auf der ersten Stufe ist. Es ist zu sehen, dass sich Formel (3.10) von Formel (3.8) nur um den Ausdruck $(1 - l/L)$ im zweiten Term unterscheidet, welcher die Endlichkeitskorrektur auf der ersten Stufe beschreibt. Bei kleinem Auswahlatz auf der ersten Stufe nähert sich diese Größe Eins an und der direkte Varianzschätzer der ersten Stufe wird fast die gesamte Varianz abdecken. Bei größerem Auswahlatz auf der ersten Stufe wird er hingegen nicht so viel Varianz auf der zweiten Stufe abdecken und eine separate Varianzschätzung auf der zweiten Stufe ist notwendig.

Ein weiterer wichtiger Faktor ist, wie viel Varianz sich auf der zweiten Stufe befindet bzw. inwieweit sich die Varianz zwischen den Stufen verteilt. Liegt ein großer Anteil der Varianz auf der ersten Stufe, wird durch Anwendung des direkten Varianzschätzers der ersten Stufe auch ein großer Anteil der gesamten Varianz geschätzt. Dies ist z.B. der Fall bei sehr homogenen Einheiten innerhalb und sehr heterogenen Einheiten zwischen den PSU. Ein anderes Anwendungsbeispiel ist eine sehr stark unterschiedliche PSU-Größe. In diesen Fällen ist die Varianz auf die ersten Stufe konzentriert. Befindet sich jedoch ein großer Anteil der Varianz auf der zweiten Stufe, z.B. bei sehr heterogenen Einheiten innerhalb der PSU, muss die zweite Stufe bei der Varianzschätzung unbedingt berücksichtigt werden. Der Varianzschätzer der ersten Stufe ist in diesem Fall nicht in der Lage, die gesamte

Varianz zu erfassen. Eine ausschließliche Berücksichtigung der ersten Stufe kann bei dieser Konstellation zu einer drastischen Unterschätzung der Varianz führen.

Der letzte entscheidende Bestimmungsfaktor ist der Auswahlatz auf der zweiten Stufe. In diesem Fall kann *ceteris paribus* davon ausgegangen werden, dass bei kleineren Stichprobenumfängen auf der zweiten Stufe der Varianzanteil durch die größere Unsicherheit zunimmt.

Aufgrund bestimmter Konstellationen dieser Parameter kann es daher notwendig sein, alle Stufen des Stichprobendesigns bei der Varianzschätzung zu berücksichtigen. Daher wird im Rahmen dieser Arbeit auf Varianzschätzer abgestellt, welche alle Stufen des Stichprobendesigns berücksichtigen.

Es wird zusätzlich noch in dieser Arbeit eine zweistufige stratifizierte Zufallsstichprobe verwendet. Die Schichtung liegt hier auf der ersten Stufe vor. Diese stellt lediglich eine Kombination der zweistufigen Zufallsstichprobe mit der geschichteten Zufallsstichprobe dar. Die Varianz und ihre Schätzung entsprechen daher denen aus (3.8) bzw. (3.9) unter Berücksichtigung einer Schichtung auf der ersten Stufe. Auf eine Angabe und Interpretation wird daher an dieser Stelle verzichtet.

Wie beschrieben gelten diese Varianzen und ihre Schätzer nur im Falle vollständiger Beobachtungen. Im Falle fehlender Werte und Imputation werden weitere Varianzkomponenten auftreten. Die dargestellten Varianzschätzer können nicht mehr angewendet werden. Dies wird in den Abschnitten 3.4.2 und 3.6 im Detail erläutert. Daher wird in den nächsten Abschnitten die Annahme vollständiger Beobachtungen aufgelöst.

3.2 Fälle der Entstehung von fehlenden Werten

Bevor es zu einer ausführlichen Erläuterung der einzelnen Komponenten der Varianz bei fehlenden Werten und Imputation kommen kann, müssen zunächst Faktoren beschrieben werden, welche den Response-Vektor z bezüglich einer Auskunft zum Untersuchungsmerkmal y determinieren. Darunter ist ein Vektor zu verstehen, welcher eine Einheit im Datensatz bezüglich dieses Merkmals als beobachtet (z.B. ausgedrückt durch Eins) oder fehlend (z.B. gekennzeichnet durch Null) klassifiziert. Dieser bestimmt, ob bestimmte Varianzkomponenten auftreten und wie die Varianz genau zerlegt wird. Weiterhin ist noch von der gezogenen Stichprobe S auszugehen mit dem Vektor s , welcher eine Einheit als für die Stichproben gezogen klassifiziert.

In diesem Zusammenhang werden in dieser Arbeit drei Fälle unterschieden. Diese werden dahingehend abgegrenzt, wo fehlende Werte entstehen und ob der Response-Vektor von Simulationsdurchlauf zu Simulationsdurchlauf neu erzeugt oder einmalig festgelegt wird. Wird zunächst der Ort der Entstehung fehlender Werte betrachtet, ergeben sich zwei Möglichkeiten. Zum einen ist es denkbar, dass die fehlenden Werte schon in der Grundgesamtheit existieren, zum anderen können sie sich als Realisationen der Stichprobe ergeben (für letzteren Fall vgl. SÄRNDAL und LUNDSTRÖM, 2005, S. 44 ff.). Fallen fehlende Werte in der Grundgesamtheit an, kann auch das zweite Abgrenzungskriterium herangezogen werden. So ist es möglich, mit einem wechselndem Response-Vektor zu arbeiten. Hier werden fehlende Werte bezüglich des Untersuchungsmerkmals in jedem Simulationsdurchlauf immer wieder neu in die Grundgesamtheit eingebaut (vgl. z.B. HAZIZA, 2010, FAY, 1991 und SHAO und STEEL, 1999). In dieser Arbeit wird in diesem Fall auch öfters von Entstehung variabler fehlender Werte oder von einem variablen Response-Vektor die Rede sein. Die Erzeugung fehlender Werte wird in diesem Fall als stochastisch aufgefasst. Zum anderen ist es denkbar, fehlende Werte bzw. den Response-Vektor einmalig in der Grundgesamtheit zu erzeugen und daraus die Stichprobenziehung wiederholt vor-

zunehmen. Die Einheiten der Grundgesamtheit besitzen für jeden Simulationsdurchlauf den gleichen Antwortstatus hinsichtlich y . Entweder sie sind immer antwortende Einheiten oder für sie liegt grundsätzlich ein fehlender Wert vor. Hier wird in der Arbeit auch von einem fixen oder festen Response-Vektor oder festen fehlenden Werten gesprochen. Der Prozess der Erzeugung fehlender Werte ist hier deterministisch. Im Vergleich zu dem Vorliegen eines konstanten Response-Vektors wird für den Fall der Möglichkeit eines variierenden Antwortstatus mehr Variation durch die wiederholte Erzeugung fehlender Werte entstehen. Dies wird im Detail in den folgenden Abschnitten sowie in der Simulationsstudie gezeigt werden. Das zweite Abgrenzungskriterium ist nur für den Fall fehlender Werte in der Grundgesamtheit relevant. Bei Annahme der Entstehung fehlender Werte in der Stichprobe werden fehlende Werte nur für die Einheiten der Stichprobe generiert. Fehlende Werte können in diesem Fall nur variabel sein, da die Stichprobe von Simulation zu Simulationsdurchlauf neu gezogen wird und hierauf basierend fehlende Werte erzeugt werden. Zusammenfassend werden die drei Fälle noch einmal in den Abbildungen 3.1, 3.2 und 3.3 dargestellt. Der Fall 1 beschreibt die Erzeugung variabler fehlender Werte in der Grundgesamtheit, der Fall 2 die Erzeugung fehlender Werte in der Stichprobe und der Fall 3 den eines fixen Response-Vektors.

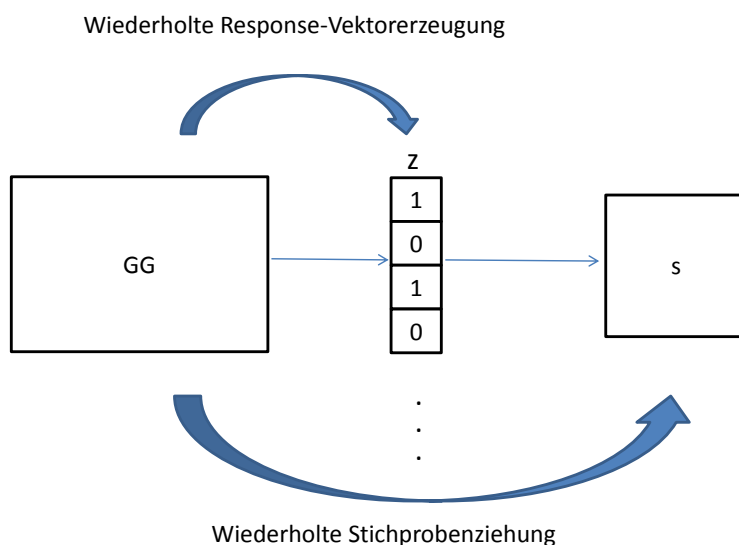


Abbildung 3.1: Fall 1: Entstehung variabler fehlender Werte in der Grundgesamtheit (vgl. hierzu auch Pfad 2 in SHAO und STEEL, 1999, S. 255)

Bezüglich der Unterscheidung von Fall 1 und Fall 2 sprechen SHAO und STEEL (1999) von zwei verschiedenen Antwortpfaden. Im Fall 1 existiert zunächst eine Grundgesamtheit, in welcher fehlende Werte vorliegen. Aus dieser Menge werden anschließend die Stichproben gezogen. In der Literatur wird die Annahme dieses Falles auch häufig *Reverse-Methode* oder *Reverse-Framework* genannt. Im Fall 2 ist es umgekehrt. Es erfolgt eine Erhebung einer Stichprobe aus einer Grundgesamtheit ohne fehlende Werte, aber bestimmte Einheiten antworten nach der Ziehung nicht. Die Annahme dieses Pfades wird in der Literatur

häufig unter den Begriffen *Zwei-Phasen-Methode* oder *Zwei-Phasen-Framework* verwendet. Fehlende Werte werden als eine Art zweite Phase der Stichprobenziehung aufgefasst (vgl. HAZIZA, 2010, S.2).

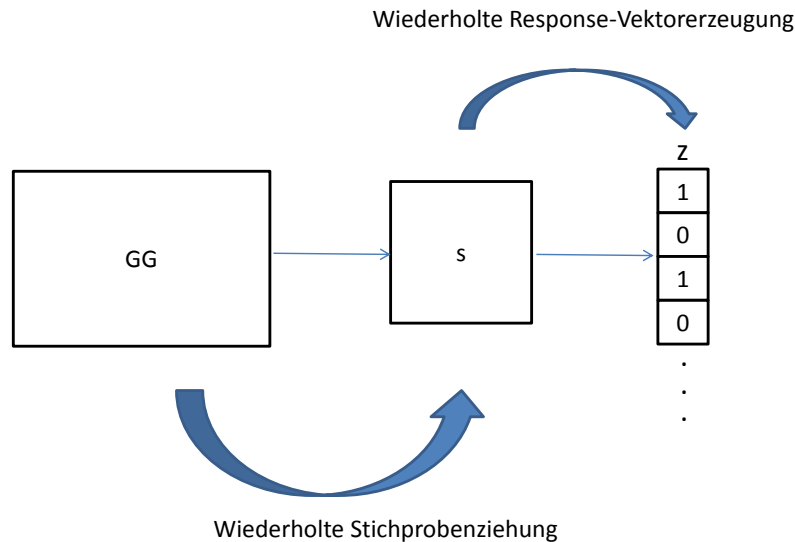


Abbildung 3.2: Fall 2: Entstehung fehlender Werte in der Stichprobe (vgl. hierzu auch Pfad 1 in SHAO und STEEL, 1999, S.255)

Unterschiede bezüglich der beiden Fällen 1 und 2 können sich insbesondere aus der Abhängigkeit der Stichprobenziehung und der Erzeugung fehlender Werte ergeben. Ein Beispiel wäre, dass die Antwortwahrscheinlichkeit einer für die Stichprobe gezogenen Einheit davon abhängig ist, welche anderen Einheiten sich in der Stichprobe befinden (vgl. BERGER et al., 2004, S. 6 f.). Solche Abhängigkeiten werden im Rahmen dieser Arbeit nicht betrachtet. Wie in BERGER et al. (2004) wird von einer Unabhängigkeit der Stichprobenziehung und der Erzeugung fehlender Werte ausgegangen. Abhängigkeiten von fehlenden Werten beziehen sich in dieser Arbeit häufig auf bestimmte Hilfsvariablen. Die Konsequenz aus der Unabhängigkeit der Stichprobenziehung und der Erzeugung fehlender Werte ist, dass die erhaltenen Stichproben- und Response-Vektoren zwischen beiden Fällen sehr ähnlich sind.

Obwohl diese Unabhängigkeit der Stichprobenziehung und der Erzeugung fehlender Werte vorausgesetzt wird, sind beide Fälle innerhalb dieser Arbeit bedeutsam. Der Grund liegt darin, dass viele Autoren zur Herleitung von Varianzen bzw. ihrer Schätzer im Falle fehlender Werte und Imputation insgesamt aber auch von einzelnen Komponenten auf den Satz der totalen Varianz zurückgreifen. Nach diesem gilt für die Varianz eines Schätzers $\hat{\theta}$:

$$V(\hat{\theta}) = E(V(\hat{\theta}|\Phi)) + V(E(\hat{\theta}|\Phi)). \quad (3.11)$$

Φ ist in diesem Fall ein Platzhalter für die bedingte Größe. Diese ergibt sich aus den beiden Fällen. Beim ersten Fall stellt der äußere Vorgang die Erzeugung fehlender Werte dar.

Der Response-Vektor wird vor der Stichprobenziehung erzeugt. Diese beschreibt den inneren Vorgang, d.h. die Stichprobenziehung erfolgt bedingt dem Response-Vektor z . Φ ist damit im ersten Fall z . Beim zweiten Fall ist es umgekehrt. Es erfolgt zunächst die Stichprobenziehung und für die gegebene Stichprobenrealisation s wird der Response-Vektor z generiert. Hier bezieht sich der innere Vorgang auf die fehlenden Werte konditioniert auf die gezogene Stichprobe. Damit wird s im zweiten Fall für Φ verwendet (vgl. hierzu auch SHAO und STEEL, 1999, S. 255). Welche Komponenten der Varianz sich für die drei Fälle genau ergeben und welche Konsequenzen für Schätzer daraus resultieren, wird in den folgenden Abschnitten und der Simulationsstudie detailliert erörtert werden.

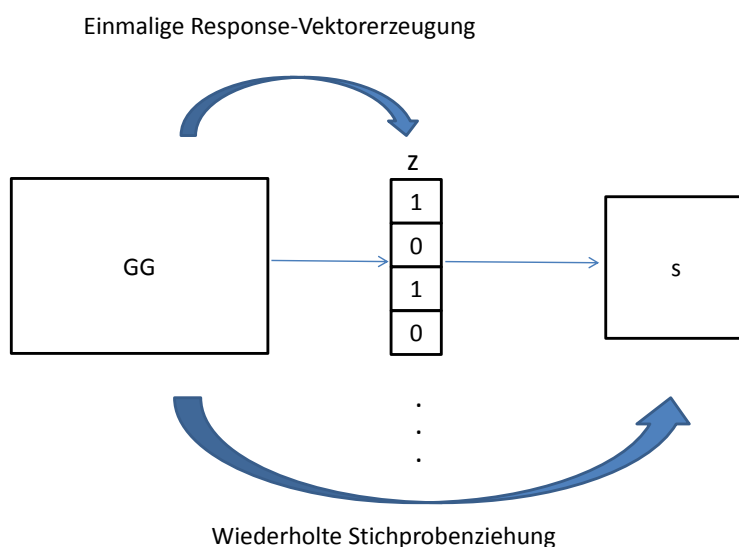


Abbildung 3.3: Fall 3: Einmalig generierter (fixer) Response-Vektor in der Grundgesamtheit

Welcher Fall in der Forschung modelliert wird, hängt meist von Faktoren wie der Präferenz des untersuchenden Wissenschaftlers und seinem Modellbildungsprozess, von der Art des herzuleitenden Varianzschätzers oder des zu analysierenden Sachverhalts, inklusive der Imputationsmethode, ab. So lassen sich z.B. bestimmte Varianzschätzer nur ableiten, wenn für die Entstehung fehlender Werte bestimmte Annahmen getroffen werden. Dies ist zum Beispiel bei den direkten Varianzschätzer in Abschnitt 3.6 öfters der Fall. Zum anderen lassen sich einige bestehende Verfahren nur analysieren, wenn hinsichtlich der Entstehung der fehlenden Werte bestimmte Annahmen getroffen werden können. So wird zum Beispiel zur Untersuchung der Erfassung bestimmter Komponenten der Varianz durch Resampling-Verfahren die Annahme der Entstehung variabler fehlender Werte in der Grundgesamtheit getroffen (vgl. HAZIZA, 2010 oder MASHREGHI et al., 2014). Im Rahmen dieser Arbeit sind, wie noch zu sehen sein wird, alle Fälle von Bedeutung und müssen jeweils in Abhängigkeit des zu analysierenden Sachverhalts herangezogen werden. Gerade der Fall eines fixen Response-Vektors ist von besonderem Interesse. So kann es in manchen Untersuchungen notwendig sein, den Prozess der Erzeugung fehlender Werte

deterministisch zu modellieren. Zudem kann dieser Fall als eine Art Vorstufe zum Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit betrachtet werden. Durch die Berücksichtigung des Falles eines fixen Response-Vektors und damit einer deterministischen Betrachtungsweise können die Auswirkungen der Behandlung fehlender Werte als stochastischer Prozess auf die Varianz und ihre Schätzung detailliert analysiert werden. In der Simulationsstudie werden aus den genannten Gründen daher alle drei Fälle berücksichtigt.

Welcher Fall in der Praxis am realistischsten ist, kann an dieser Stelle nicht geklärt werden. Alle Fälle sind denkbar. Eine Person kann z.B. grundsätzlich immer die Auskunft verweigern, ihr Antwortverhalten kann aber auch variieren oder von der konkreten Erhebung abhängen.³

3.3 Varianz bei fehlenden Werten

Ziel des folgenden Abschnittes ist die allgemeinen Auswirkungen fehlender Werte auf die Varianz des Schätzers $\hat{\theta}_{\text{NR}}$ zu untersuchen. Dabei wird auf die Ausführungen von SÄRNDAL und LUNDSTRÖM (2005, S. 44 ff.) zurückgegriffen. Diese gehen von der Entstehung fehlender Werte in der Stichprobe aus, wie in Abbildung 3.2 dargestellt.

In diesem Abschnitt wird noch nicht auf die Imputation und deren Auswirkungen auf die Varianz eingegangen. Daher wird der Schätzer in diesem Abschnitt statt $\hat{\theta}_I$ mit $\hat{\theta}_{\text{NR}}$ bezeichnet, um den Fokus der Auswirkungen fehlender Werte (*Nonresponse*: NR) hervorzuheben.

Um die Auswirkungen von fehlenden Werten auf die Varianz eines Schätzers zu verstehen, gilt es zunächst, die Auswirkungen auf den entsprechenden Punktschätzer $\hat{\theta}_{\text{NR}}$ zu betrachten. Durch das Auftreten von fehlenden Werten wird die gesamte Verzerrung, welche durch die Schätzung $\hat{\theta}_{\text{NR}} - \theta$ entsteht, bei Ausblendung der Imputation, aufgesplittet in:

$$\hat{\theta}_{\text{NR}} - \theta = (\hat{\theta} - \theta) + (\hat{\theta}_{\text{NR}} - \hat{\theta}). \quad (3.12)$$

$\hat{\theta} - \theta$ beschreibt den Stichprobenfehler, also den Fehler, der aus der Betrachtung der Stichprobe an der Stelle der gesamten Bevölkerung entsteht. Der zweite Term $\hat{\theta}_{\text{NR}} - \hat{\theta}$ beschreibt den Nonresponse-Fehler, also jenen Fehler, der aufgrund fehlender Werte in der Stichprobe auftritt (vgl. SÄRNDAL und LUNDSTRÖM, 2005, S. 44 ff.).

Wie beschrieben, gehen SÄRNDAL und LUNDSTRÖM (2005) von einer Generierung fehlender Werte in der Stichprobe aus, wie in Abbildung 3.2 dargestellt. Es werden zunächst alle möglichen Stichproben S gebildet, welche bezüglich eines bestimmten Stichprobendesigns gezogen werden können, beschrieben durch $p(S)$. Auf der zweiten Stufe werden alle möglichen Response-Vektoren z berechnet, welche für einen gegebenen Stichprobenvektor s für eine unbekannte Response-Verteilung realisiert werden können, beschrieben durch $q(z|s)$. Die entsprechenden Erwartungswerte und Varianzen bezüglich beider Verteilungen werden mit E_p bzw. V_p und E_q bzw. V_q bezeichnet. Operationen, welche sich auf beide Verteilungen gemeinsam beziehen, werden mit Index pq gekennzeichnet.

³ Vgl. zu den Ausführungen in diesem Abschnitt auch die Untersuchungen im Dacseis-Projekt (<https://www.uni-trier.de/index.php?id=29730>), insbesondere die Berichte von BERGER et al. (2004) und BIHLER et al., 2004, sowie die Veröffentlichung von BJØRNSTAD, 2007 mit dazugehöriger Diskussion.

Während der Stichprobenfehler $\hat{\theta} - \theta$ bei geeigneter Designwahl meist Null oder Nahe an Null ist, wird der Nonresponse-Fehler $\hat{\theta}_{\text{NR}} - \hat{\theta}$ häufig von Null abweichen. Der Grund liegt darin, dass fehlende Werte in Abhängigkeit der beschriebenen Ausfallmechanismen zu einer gewissen Verzerrung führen. Der bedingte erwartete Nonresponse-Fehler, also die Nonresponse-Verzerrung bei einer gegebenen Stichprobe, geben SÄRNDAL und LUNDSTRÖM (2005) an durch:

$$B_{\text{NR}|s} = E_q((\hat{\theta}_{\text{NR}} - \hat{\theta})|s) = E_q(\hat{\theta}_{\text{NR}}|s) - \hat{\theta}. \quad (3.13)$$

Die gesamte Nonresponse-Verzerrung wird errechnet durch Bildung des Erwartungswertes über alle möglichen Stichproben:

$$B_{\text{NR}} = E_p(B_{\text{NR}|s}) = E_{pq}(\hat{\theta}_{\text{NR}} - \hat{\theta}) = E_{pq}(\hat{\theta}_{\text{NR}}) - E_p(\hat{\theta}). \quad (3.14)$$

Der erwartete Stichprobenfehler ist weiterhin gegeben durch:

$$B_{\text{SAM}} = E_p(\hat{\theta} - \theta) = E_p(\hat{\theta}) - \theta. \quad (3.15)$$

Die gesamte Verzerrung der interessierenden Statistik $\hat{\theta}_{\text{NR}}$ ist damit beschrieben durch:

$$B_{pq}(\hat{\theta}_{\text{NR}}) = E_{pq}(\hat{\theta}_{\text{NR}} - \theta) = B_{\text{SAM}} + B_{\text{NR}}. \quad (3.16)$$

Damit setzt sich die gesamte Verzerrung aus einer Stichproben- und einer Nonresponse-Komponente zusammen.

Neben der gesamten Verzerrung ist es auch möglich die Varianz des Schätzers $\hat{\theta}_{\text{NR}}$ zu zerlegen. Allgemein wird diese angegeben durch:

$$V(\hat{\theta}_{\text{NR}}) = E_{pq}(\hat{\theta}_{\text{NR}} - E_{pq}(\hat{\theta}_{\text{NR}}))^2. \quad (3.17)$$

Um diese Varianz präziser darzustellen, bedienen sich SÄRNDAL und LUNDSTRÖM (2005) der Varianzzerlegung aus Abschnitt 3.2, gemäß derer die Varianz eines Schätzers die Varianz der konditionierten Erwartung plus die Erwartung der bedingten Varianz ist. Sie nutzen hierbei aus, dass gemäß des zweiten Pfades auf die Stichprobe konditioniert wird. Es gilt:

$$V(\hat{\theta}_{\text{NR}}) = V_p E_q(\hat{\theta}_{\text{NR}}|s) + E_p V_q(\hat{\theta}_{\text{NR}}|s). \quad (3.18)$$

Der erste Term bezieht sich auch hier auf die Stichprobenziehung, insbesondere die dadurch entstehende Variabilität und der zweite Term auf die Variabilität durch fehlende Werte. Eine genauere Interpretation der einzelnen Komponenten wird in Zusammenhang mit der Imputation in Abschnitt 3.4.2 gegeben. An dieser Stelle wird eher eine formale Auseinandersetzung stattfinden mit dem Ziel der Darstellung der Auswirkung fehlender Werte auf die Varianz. Nach (3.13) ist $E_q(\hat{\theta}_{\text{NR}}|s) = \hat{\theta} + B_{\text{NR}|s}$ und damit:

$$V(\hat{\theta}_{\text{NR}}) = V_p(\hat{\theta} + B_{\text{NR}|s}) + E_p V_q(\hat{\theta}_{\text{NR}}|s). \quad (3.19)$$

Kann aufgrund einer bestimmten Verteilung $q(z|s)$ die Annahme getroffen werden, dass die bedingte Verzerrung durch fehlende Werte $B_{\text{NR}|s} = 0$ ist, vereinfacht sich die Varianz (3.19) zu

$$V(\hat{\theta}_{\text{NR}}) = V_p(\hat{\theta}) + E_p V_q(\hat{\theta}_{\text{NR}}|s). \quad (3.20)$$

Der erste Term $V_p(\hat{\theta})$ beschreibt in diesem Fall die Stichprobenvarianz, wie sie im Falle vollständiger Beobachtungen vorliegt (vgl. Abschnitt 3.1). Sie hängt nicht von fehlenden Werten oder der Verteilung $q(z|s)$ ab. $E_p V_q(\hat{\theta}_{\text{NR}}|s)$ beschreibt weiterhin die Varianz, welche durch die fehlenden Werte entsteht.

Gleichung 3.20 zeigt damit, dass, selbst wenn keine Verzerrung des Punktschätzers aufgrund der fehlenden Werte vorliegt, diese dennoch zu einer Steigerung der Varianz führen kann. Dies ergibt sich im Vergleich zur Varianz des Punktschätzers im Falle einer vollständigen Beobachtung wie in Abschnitt 3.1 definiert. Diese beträgt lediglich $V_p(\hat{\theta})$, wird also nur durch die Stichprobenvarianz determiniert. Liegt jedoch eine bedingte Verzerrung durch fehlende Werte vor, damit $B_{\text{NR}|s} \neq 0$, gilt (3.19). Gemäß der Varianz zweier Zufallsvariablen kann $V_p(\hat{\theta} + B_{\text{NR}|s})$ beschrieben werden durch:

$$V_p(\hat{\theta} + B_{\text{NR}|s}) = V_p(\hat{\theta}) + V_p(B_{\text{NR}|s}) + 2 \cdot \text{Cov}_p(\hat{\theta}, B_{\text{NR}|s}).$$

Für die Varianz des Punktschätzers $V(\hat{\theta}_{\text{NR}})$ aus (3.18) resultiert:

$$\begin{aligned} V(\hat{\theta}_{\text{NR}}) &= V_p(\hat{\theta}) + E_p V_q(\hat{\theta}_{\text{NR}}|s) \\ &\quad + V_p(B_{\text{NR}|s}) + 2 \cdot \text{Cov}_p(\hat{\theta}, B_{\text{NR}|s}). \end{aligned} \quad (3.21)$$

Gegenüber (3.20) ergibt sich damit noch eine zusätzliche Varianzerhöhung um den Summanden $V_p(B_{\text{NR}|s}) + 2 \cdot \text{Cov}_p(\hat{\theta}, B_{\text{NR}|s})$ wenn zusätzlich eine Verzerrung $B_{\text{NR}|s} \neq 0$ aufgrund von fehlenden Werten und einer bestimmten Verteilung $q(z|s)$ vorliegt (vgl. SÄRNDAL und LUNDSTRÖM, 2005, S. 47). Es kann also festgehalten werden, dass fehlende Werte zu einer Erhöhung der Varianz eines Punktschätzers führen. Es müssen Varianzschätzer gefunden werden, welche die verschiedenen Komponenten abdecken. Jedoch sind die in diesem Abschnitt beschriebenen Komponenten noch nicht ausreichend. Maßnahmen zur Kompensation fehlender Werte besitzen ebenfalls Einfluss auf die Varianz eines Schätzers.

3.4 Varianz bei Imputation

Bisher wurde nur allgemein auf die Auswirkungen fehlender Werte auf die Varianz eines Schätzers eingegangen. Die Auswirkungen der Anwendung bestimmter Maßnahmen zur Kompensation fehlender Werte auf die Varianz eines Schätzers wurde bis jetzt unterschlagen. In diesem Abschnitt wird insbesondere untersucht, wie die Anwendung der in Abschnitt 2.2 beschriebenen Imputationsverfahren die Varianz eines Schätzers beeinflusst. In diesem Zusammenhang ist die Zerlegung der Gesamtvarianz in ihre einzelnen Komponenten von Bedeutung und hier spielen u.a. die drei Fälle der Entstehung fehlender Werte eine Rolle.

3.4.1 Quellen der Unsicherheit und die dazugehörenden Varianzkomponenten bei Imputation hinsichtlich der Entstehung fehlender Werte

Diese drei Fälle der Entstehung fehlender Werte wurden in Abschnitt 3.2 ausführlich erläutert (vgl. zu den Ausführungen dieses Abschnittes auch SHAO und STEEL, 1999,

FAY, 1991, MASHREGHI et al., 2014, HAZIZA, 2010, CHAUVET et al., 2011, SÄRNDAL und LUNDSTRÖM, 2005, BERGER et al., 2004, BJØRNSTAD, 2007 mit dazugehöriger Diskussion sowie BIHLER et al., 2004). Im Rahmen der drei Fälle entstehen verschiedene Quellen der Unsicherheit, welche sich insbesondere aus dem Zusammenspiel von Stichprobenziehung, dem Auftreten fehlender Werte und der Imputation ergeben. Aus diesen Quellen der Unsicherheit resultieren die einzelnen Varianzkomponenten der Gesamtvarianz. Die drei beschriebenen Fälle der Entstehung fehlender Werte bestimmen daher sowohl das Auftreten wie auch die Reihenfolgen des Auftretens der einzelnen Komponenten.

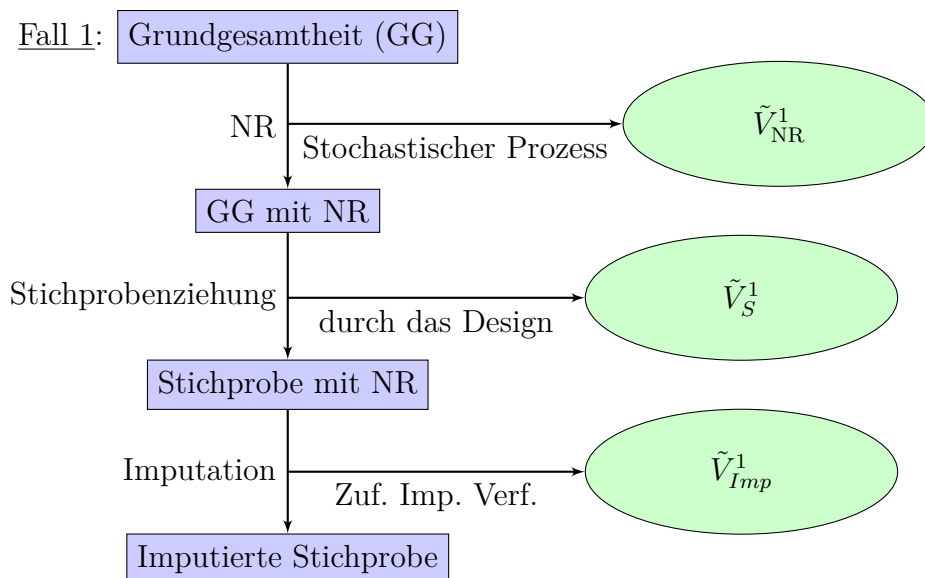


Abbildung 3.4: Quellen der Unsicherheit mit dazugehörigen Varianzkomponenten in Fall 1 und der Verwendung eines zufälligen Imputationsverfahrens (vgl. hierzu auch Pfad 2 in SHAO und STEEL, 1999, S. 255)

Der erste Fall wurde in Abschnitt 3.2 mit der Entstehung variabler fehlender Werte in der Grundgesamtheit bezeichnet. Der dazugehörige Prozess wird in Abbildung 3.4 dargestellt. Wie in Abschnitt 3.2 erläutert, wird hier zunächst von einer Grundgesamtheit ausgegangen, in welcher fehlende Werte entstehen. In diesem Fall wird dieser Vorgang als stochastischer Prozess aufgefasst. Es wird damit unterstellt, dass das Antwortverhalten nicht grundsätzlich festgelegt ist und sich von Fall zu Fall ändern kann. Im Rahmen der Simulationsstudie wird in jedem Simulationsdurchlauf die Generierung des Response-Vektors neu durchgeführt. Durch diesen Vorgang entsteht die erste Quelle der Unsicherheit bzw. Variabilität, welche in Zusammenhang mit der Entstehung fehlender Werte steht. Die entsprechende Varianzkomponente wird als Varianzkomponente bezüglich fehlender Werte bezeichnet und durch \tilde{V}_{NR}^1 gekennzeichnet. Die Angabe im Exponent bezieht sich auf den Fall der Entstehung von fehlenden Werten und damit hier auf den ersten Fall. Die Tilde über den einzelnen Varianztermen soll zum Ausdruck bringen, dass die einzelnen Komponenten bezüglich eines zufälligen Imputationsverfahrens zu betrachten sind.⁴

Aufgrund des Auftretens fehlender Werte liegt nun eine Grundgesamtheit mit fehlenden Werten vor. Aus dieser wird anschließend eine Stichprobe nach einem bestimmten Design gezogen. Dadurch ergibt sich die nächste Quelle der Unsicherheit und zwar diejenige die mit der Stichprobenziehung zusammenhängt. Diese wird als Stichprobenkomponente (der gesamten Varianz des imputierten Schätzers) bezeichnet, angegeben durch den Term \tilde{V}_S^1 .

⁴ Im Falle eines deterministischen Imputationsverfahrens werden die Varianzterme ohne Tilde angegeben.

Die Variabilität wird naturgemäß insbesondere vom zugrunde liegenden Stichprobendesign determiniert.

Aufgrund der Stichprobenziehung resultiert eine Stichprobe mit fehlenden Werten. Wird zur Kompensation der fehlenden Werte die Imputation gewählt und dabei auf ein zufälliges Imputationsverfahren zurückgegriffen, entsteht die letzte Quelle der Unsicherheit. Wie in Abschnitt 2.2.1 beschrieben, führen zufällige Imputationsverfahren zu größeren Varianzen eines bestimmten Punktschätzers als deterministische Imputationsverfahren. Bei zufälligen Methoden entsteht zusätzliche Variabilität durch die randomisierte Auswahl von Imputationswerten bzw. deren Komponenten, wie z.B. zufälliger Störgrößen. Die dadurch auftretende Varianz wird als Imputationsvarianz bezeichnet. Die entsprechende Varianzkomponente wird als Imputationskomponente der Varianz definiert und durch das Symbol \tilde{V}_{Imp}^1 ausgedrückt.

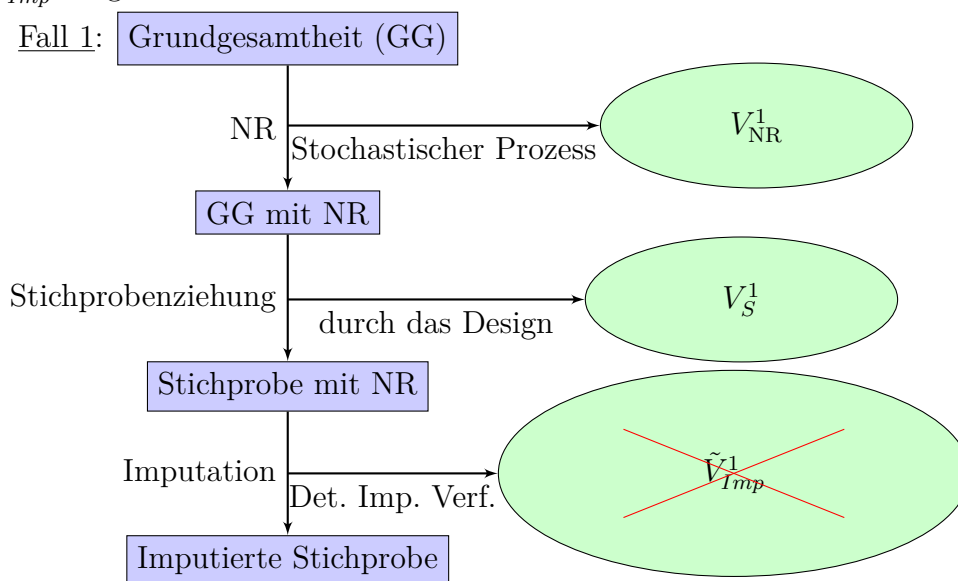


Abbildung 3.5: Quellen der Unsicherheit mit dazugehörigen Varianzkomponenten in Fall 1 und der Verwendung eines deterministischen Imputationsverfahrens (vgl. hierzu auch Pfad 2 in SHAO und STEEL, 1999, S. 255)

Die Imputationskomponente fällt nicht an, wenn ein deterministisches Imputationsverfahren verwendet wird, wie in Abbildung 3.5 zu sehen ist.⁵

Wie in Abschnitt 3.2 beschrieben, ändert sich in Fall 2 der Entstehungsort fehlender Werte. Diese entstehen zu einem späteren Zeitpunkt in der Stichprobe, wie in Abbildung 3.6 dargestellt. Hierdurch ändert sich die Reihenfolge, des Auftretens der Quellen der Unsicherheit. Aus der Grundgesamtheit erfolgt zunächst die Stichprobenziehung, wodurch zunächst die Stichprobenkomponente der Varianz \tilde{V}_S^2 betroffen ist. Fehlende Werte entstehen anschließend in der Stichprobe, wobei erst an dieser Stelle die Varianzkomponente bezüglich fehlender Werte tangiert wird.

In Abschnitt 3.2 wurde bereits erläutert, dass der äußere Vorgang hier die Erzeugung fehlender Werte darstellt und sich der innere Vorgang auf die Stichprobenziehung bezieht. Bei Fall 1 ist es genau umgekehrt. Der Imputationsvorgang ist in beiden Fällen immer der innerste Vorgang, da die Imputationen auf die fehlenden Werte in einer Stichprobe vorgenommen werden.

⁵ Die Symbole der Stichprobenkomponente und der Varianzkomponente bezüglich fehlender Werte besitzen hier keine Tilde, da sich die beiden Varianzkomponenten, wie zuvor beschrieben, im Vergleich zum Fall eines zufälligen Imputationsverfahren unterscheiden.

Wie im weiteren Verlauf der Arbeit noch zu sehen sein wird, führt diese unterschiedliche Reihenfolge zu einer veränderten Zusammensetzung der Gesamtvarianz des imputierten Schätzers. Aufgrund der in Abschnitt 3.2 getroffenen Annahme der Unabhängigkeit der Stichprobenziehung und der Erzeugung fehlender Werte resultiert jedoch eine gleiche Gesamtvarianz. Auch dieser Sachverhalt wird im weiteren Verlauf der Arbeit zu sehen sein.

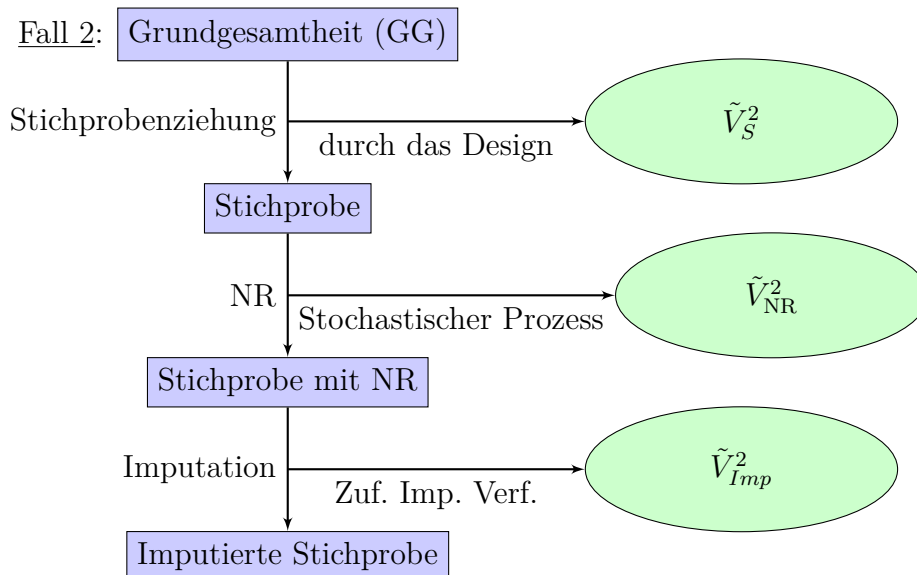


Abbildung 3.6: Quellen der Unsicherheit mit dazugehörigen Varianzkomponenten in Fall 2 und der Verwendung eines zufälligen Imputationsverfahrens (vgl. hierzu auch Pfad 1 in SHAO und STEEL, 1999, S. 255)

Der dritte und damit letzte Fall der Entstehung fehlender Werte ist in Abbildung 3.7 angegeben. Dieser Fall wurde in Abschnitt 3.2 als Fall eines fixen Response-Vektors bezeichnet. Wie im ersten Fall wird auch hier von der Entstehung fehlender Werte in der Grundgesamtheit ausgegangen. Jedoch wird dieser Prozess der Generierung fehlender Werte als deterministisch aufgefasst. Damit wird den Befragten unterstellt, dass sie ein unveränderliches Antwortverhalten besitzen. In der Simulation wird dadurch in jedem Simulationsdurchlauf mit dem gleichen Response-Vektor gearbeitet. Aus diesem Grund kann die Variation, welche mit einem in jedem Simulationsdurchlauf neu generierten Response-Vektor einhergeht, nicht entstehen. Die Varianzkomponente bezüglich fehlender Werte \tilde{V}_{NR}^1 tritt im Vergleich zu Fall 1 hier nicht auf. Für die beiden anderen Komponenten gelten jedoch die in diesem Fall getätigten Erläuterungen.

Hierdurch stellt sich die Frage, inwieweit die drei beschriebenen Quellen der Unsicherheit in der Praxis beobachtbar sind. Das Ziel besteht darin, die dazugehörigen Varianzkomponenten über die Varianzschätzung zu messen. Als problematisch erweist sich hierbei die Varianzkomponente bezüglich fehlender Werte dar. Wie in den Abschnitten 2.1 beschrieben, ist die Response-Verteilung bzw. der Ausfallmechanismus in den meisten Fällen unbekannt (vgl. hierzu auch SÄRNDAL, 1992, S. 245). Informationen über das Antwortverhalten sind häufig sehr begrenzt. Damit wird die Varianzkomponente bezüglich fehlender Werte in der Praxis kaum beobachtbar sein. Eine Möglichkeit, diese Komponente messbar zu machen, ist die Verwendung von Simulationen. Hier können Annahmen über den Ausfallmechanismus bzw. das Antwortverhalten getroffen und die Auswirkungen auf die Schätzer untersucht werden. Es können auch die Auswirkungen verschiedener Annahmen verglichen werden. Auch im Rahmen dieser Arbeit ist es das Ziel die Varianzkomponente bezüglich fehlender Werte über eine Simulationsstudie messbar zu machen (vgl. Kapitel

5).

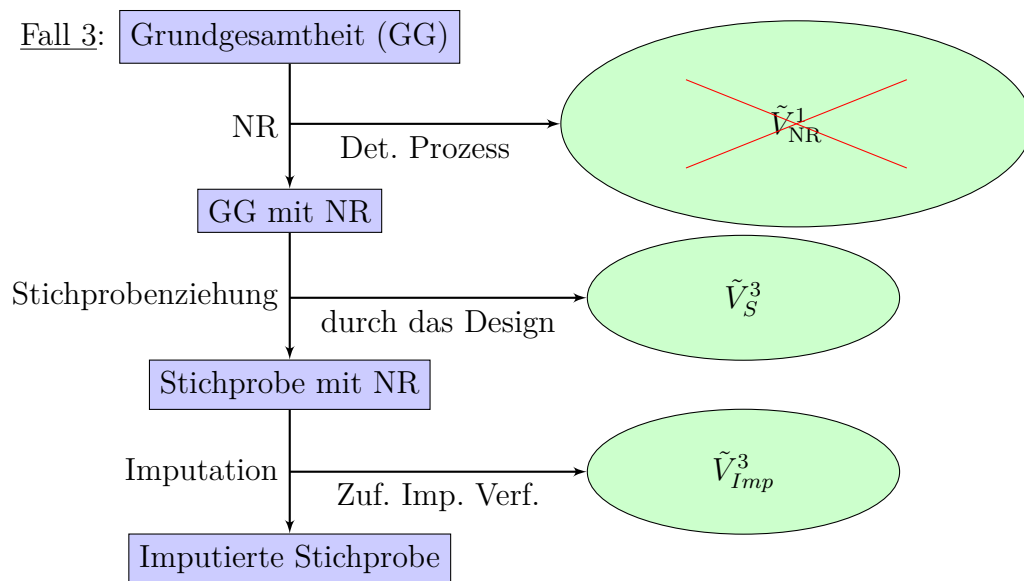


Abbildung 3.7: Quellen der Unsicherheit mit dazugehörigen Varianzkomponenten in Fall 3 und der Verwendung eines zufälligen Imputationsverfahrens

Die beiden anderen Komponenten sind hingegen unkritischer. Zumindest für den Datenbereitsteller sollte klar sein, nach welchem Stichprobendesign die Ziehung vorgenommen und nach welcher Methode die Imputation durchgeführt worden ist (vgl. hierzu auch SÄRNDAL, 1992, S. 245). Daher werden diese beiden Komponenten in der Praxis einfacher zu beobachten sein.

3.4.2 Formale Zerlegung der Gesamtvarianz in einzelne Varianzkomponenten

Zur Ableitung geeigneter Varianzschätzer ist eine formale Zerlegung der Varianz notwendig, d.h. die Gesamtvarianz eines Schätzers wird in die zuvor beschriebenen Varianzkomponenten zerlegt.

Die formale Betrachtungsweise erfordert zunächst eine Untersuchung des Gesamtfehlers $\hat{\theta}_I - \theta$, welcher gemäß CHAUVET et al. (2011) bei zufälligen Imputationsverfahren ausgedrückt werden kann durch:

$$\hat{\theta}_I - \theta = (\hat{\theta} - \theta) + (E_I(\hat{\theta}_I) - \hat{\theta}) + (\hat{\theta}_I - E_I(\hat{\theta}_I)). \quad (3.22)$$

E_I bezeichnet den Erwartungswert mit Bezug auf den Imputationsmechanismus, welcher verwendet wird um den zufälligen Term zu ziehen. Der zweite Term aus (3.22) beschreibt den Nonresponse-Fehler und der dritte Term den Imputationsfehler. Der erste Term ist wie zuvor der Stichprobenfehler. Im Falle deterministischer Imputationsverfahren wäre der Imputationsfehler Null (vgl. CHAUVET et al., 2011, S. 462).

Zur Zerlegung der Gesamtvarianz des Schätzers $\hat{\theta}_I$ sind auch bei der formalen Betrachtung die drei Fälle von fehlenden Werten interessant. Je nachdem ob hier ein deterministisches oder zufälliges Imputationsverfahren vorliegt, ob von einem festen oder variablen Response-Vektor ausgegangen wird und nach welchen Größen die inneren Vorgänge im

Sinne von SHAO und STEEL (1999) konditioniert werden, ergeben sich verschiedene Varianzzerlegungen.

Zunächst wird der Fall deterministischer Imputationsverfahren betrachtet. MASHREGHI et al. (2014, S.145) geben hier die Varianz des imputierten Schätzers $\hat{\theta}_I$ im Falle der Entstehung variabler fehlender Werte in der Grundgesamtheit an, wobei sich die Indizes p und q wieder auf das Stichprobendesign bzw. den Antwortmechanismus beziehen:

$$V(\hat{\theta}_I) = \underbrace{E_q V_p(\hat{\theta}_I|z)}_{V_S^1} + \underbrace{V_q E_p(\hat{\theta}_I|z)}_{V_{NR}^1}. \quad (3.23)$$

Die Gesamtvarianz besteht hier aus zwei Komponenten, welche wiederum bedingte Erwartungswerte und Varianzen beinhalten. Da hier von variablen fehlenden Werten in der Grundgesamtheit ausgegangen wird, erfolgt eine Konditionierung auf den Response-Vektor z . Die erste Komponente ist wie folgt zu deuten. Der innere Vorgang bezieht sich auf das Stichprobendesign. Er beschreibt die Varianz des Schätzers $\hat{\theta}_I$ über alle möglichen Stichproben, welche für einen gegebenen Response-Vektor z gebildet werden können. Anschließend erfolgt die Erwartungswertbildung über alle möglichen Response-Vektoren z . Entscheidend ist jedoch, dass die Varianz dieser Komponente über die Stichprobenziehung entsteht. Daher wird diese erste Komponente im Rahmen der Arbeit, wie in Abschnitt 3.4.1 erläutert, als Stichprobenkomponente (der gesamten Varianz des imputierten Schätzers) V_S^1 bezeichnet. Im Gegensatz hierzu wird innerhalb der zweiten Komponente aus (3.23) die Variabilität über die fehlenden Werte generiert. Hier wird zunächst der Erwartungswert des Schätzers $\hat{\theta}_I$ über die möglichen Stichproben für einen gegebenen Response-Vektor z berechnet. Die Varianz V_q bezieht sich auf alle möglichen Response-Vektoren. Daher wird dieser Term, wie in Abschnitt 3.4.1 beschrieben, im Rahmen der Arbeit auch als Varianzkomponente bezüglich fehlender Werte V_{NR}^1 bezeichnet.

Aber auch für die beiden anderen Fälle der Entstehung fehlender Werte können die Varianzen formal definiert werden. Entstehen die fehlenden Werte in der Stichprobe, gilt für die Varianz eines Schätzers bei deterministischen Imputationsverfahren:

$$V(\hat{\theta}_I) = \underbrace{V_p E_q(\hat{\theta}_I|s)}_{V_S^2} + \underbrace{E_p V_q(\hat{\theta}_I|s)}_{V_{NR}^2}. \quad (3.24)$$

Dies ist der Formel (3.18) sehr ähnlich, wobei hier vom konkreten Fall der Imputation ausgegangen wird.

Der innere und äußere Vorgang sind im Vergleich zum Fall variabler fehlender Werte in der Grundgesamtheit gegenteilig zu betrachten. Dieser Sachverhalt wurde schon in den Abschnitten 3.2 und 3.4.1 ausführlich diskutiert. Damit wird für die Stichprobenkomponente V_S^2 zunächst der Erwartungswert des Schätzers $\hat{\theta}_I$ über alle möglichen Response-Vektoren gegeben einer konkreten Stichprobenrealisation berechnet. Anschließend erfolgt die Varianzbildung über alle möglichen Stichproben. Im Rahmen der Komponente V_{NR}^2 wird zunächst die Varianz des Schätzers $\hat{\theta}_I$ über alle möglichen Response-Vektoren für die Stichprobenrealisation s kalkuliert. Anschließend erfolgt die Erwartungswertbildung über alle möglichen Stichproben.

Bei fixierten fehlenden Werten in der Grundgesamtheit tritt die Varianzkomponente bezüglich fehlender Werte nicht auf, da die Erzeugung des Response-Vektors z einmalig erfolgt. Dieser Prozess ist deterministisch. Der Vektor z stellt in diesem Fall keine Zufallsvariable, sondern eine Konstante dar. Der äußere Vorgang der Erwartungswertbildung bezüglich

des Antwortmechanismus E_q muss damit nicht betrachtet werden. Nur die Variabilität des Schätzers $\hat{\theta}_I$ bezüglich der Stichprobenziehung für den gegebenen Response-Vektor z tritt damit in Erscheinung. Daher wird (3.23) modifiziert zu:

$$V(\hat{\theta}_I) = \underbrace{V_p(\hat{\theta}_I|z)}_{V_S^3}. \quad (3.25)$$

Für zufällige Imputationsverfahren geben MASHREGHI et al. (2014, S.145) die Varianz eines Schätzers bei variablen fehlenden Werten in der Grundgesamtheit an durch:

$$V(\hat{\theta}_I) = \underbrace{E_q V_p E_I(\hat{\theta}_I|s, z)}_{\tilde{V}_S^1} + \underbrace{V_q E_p E_I(\hat{\theta}_I|s, z)}_{\tilde{V}_{NR}^1} + \underbrace{E_q E_p V_I(\hat{\theta}_I|s, z)}_{\tilde{V}_{Imp}^1}. \quad (3.26)$$

Entstehen fehlende Werte hingegen in der Stichprobe kann (3.26) umgeformt werden zu:

$$V(\hat{\theta}_I) = \underbrace{V_p E_q E_I(\hat{\theta}_I|s, z)}_{\tilde{V}_S^2} + \underbrace{E_p V_q E_I(\hat{\theta}_I|s, z)}_{\tilde{V}_{NR}^2} + \underbrace{E_p E_q V_I(\hat{\theta}_I|s, z)}_{\tilde{V}_{Imp}^2}, \quad (3.27)$$

(vgl. CHAUVET et al., 2011).

Sind die fehlenden Werte in der Grundgesamtheit fixiert, ist die Varianz über die verschiedenen Response-Vektoren V_q null und der mittlere Term aus (3.26) entfällt. Da z in diesem Fall eine Konstante darstellt, braucht auch hier der Erwartungswert E_q nicht betrachtet zu werden. Es resultiert:

$$V(\hat{\theta}_I) = \underbrace{V_p E_I(\hat{\theta}_I|s, z)}_{\tilde{V}_S^3} + \underbrace{E_p V_I(\hat{\theta}_I|s, z)}_{\tilde{V}_{Imp}^3}. \quad (3.28)$$

Bei zufälligen Imputationsverfahren kommt eine dritte Varianzkomponente hinzu, welche aus der Zufälligkeit der Auswahl von Imputationswerten resultiert. Der letzte Term in allen drei Gleichungen stellt diese Varianzkomponente dar.⁶ Sie wird, wie in Abschnitt 3.4.1 dargelegt, als Imputationskomponente bezeichnet, abgekürzt durch \tilde{V}_{Imp}^* , wobei * den jeweiligen Fall fehlender Werte bezeichnet. V_I beschreibt die Varianz, welche in Zusammenhang mit der Imputation steht. Diese Erwartungswert- bzw. Varianzberechnung bezüglich der Imputation bildet in allen Komponenten der Varianz den innersten Vorgang, unabhängig, welcher der Fälle (3.26),(3.27) oder (3.28) vorliegt. Wie in Abschnitt 3.4.1 erläutert, liegt der Grund darin, dass die Berechnung der Imputationswerte auf Basis der Stichprobenziehung und des Response-Vektors erfolgt. Imputationswerte können naturgemäß erst dann berechnet werden, wenn klar ist, für welche Einheiten fehlende Werte vorliegen bzw. welche Einheiten überhaupt in die Stichprobe gelangen. $E_I(\hat{\theta}_I|s, z)$ und $V_I(\hat{\theta}_I|s, z)$ sind dahingehend so zu interpretieren, dass für eine gegebene Stichprobenrealisation s und einen Response-Vektor z die Erwartungswert- und Varianzbildung bezüglich des Schätzers $\hat{\theta}_I$ über alle möglichen Imputationswerte erfolgt. Die Erwartungswerte und

⁶ Es wird hier grundsätzlich von der dritten Komponente gesprochen, auch wenn im Fall eines fixen Response-Vektors nur zwei Komponenten vorhanden sind. Dies geschieht aus Gründen der Einheitlichkeit. Die Komponente bezüglich fehlender Werte stellt die zweite Komponente dar und wird in diesem Fall als nicht auftretend und damit null interpretiert.

Varianzen E_p , V_p , E_q und V_q sind anschließend in Abhängigkeit des Falls an fehlenden Werten und der zu berechneten Komponente über diesen innersten Vorgang zu bilden. Die Gleichung (3.26) lässt sich einfach zur Gleichung (3.23) überführen. Im deterministischen Fall ist $V_I(\hat{\theta}_I|s, z)$ gleich Null und damit der gesamte dritte Term, die Imputationskomponente. Gegeben s und z ergibt sich bei deterministischen Imputationsverfahren eine einzig mögliche Realisation an Imputationswerten. Daher braucht hier die Erwartungswertbildung E_I nicht betrachtet zu werden. Gleiche Erläuterungen können für die beiden anderen Fälle von fehlenden Werten vollzogen werden.

3.4.3 Beitrag der einzelnen Komponenten zur Gesamtvarianz

Nach HAZIZA (2010) und MASHREGHI et al. (2014) tragen die einzelnen Varianzkomponenten in Abhängigkeit bestimmter Parameter in einem unterschiedlichen Ausmaß zur Gesamtvarianz eines Schätzers bei. Dies wird erörtert unter der Annahme der Entstehung fehlender Werte in der Grundgesamtheit. HAZIZA (2010) und MASHREGHI et al. (2014) geben die Ordnung des ersten Terms aus (3.23), die Stichprobenkomponente V_S^1 , mit $\mathcal{O}(1/n)$ an. Die Varianzkomponente bezüglich fehlender Werte V_{NR}^1 ist jedoch nur von Ordnung $\mathcal{O}(1/N)$. Der Beitrag zur gesamten Varianz des zweiten Terms beziffern sie mit $\mathcal{O}(n/N)$ und ist nur signifikant bei einem großen Gesamtauswahlsatz $f = \frac{n}{N}$. Intuitiv kann dies damit erklärt werden, dass bei größer werdendem Stichprobenumfang n die Variabilität durch die Stichprobenziehung geringer wird. Die Vernachlässigbarkeit von V_{NR}^1 steht nach MASHREGHI et al. (2014, S. 145) z.B. auch bei der geschichteten Zufallsstichprobe im Zusammenhang mit dem Gesamtauswahlsatz f . Die Auswahlsätze in den einzelnen Schichten f_h müssen hier nicht geringfügig sein.

Die beschriebene Abhängigkeit der Varianzkomponenten in ihrer absoluten Höhe wie auch ihres relativen Beitrags zur Gesamtvarianz von bestimmten Parametern wird im Folgenden an zwei konkreten Imputationsmethoden demonstriert. Zunächst erfolgt die Betrachtung der Mittelwertimputation bei Verwendung der einfachen Zufallsstichprobe und des Mittelwertschätzers. Nach MASHREGHI et al. (2014, S. 150) kann unter bestimmten Annahmen, wie dem uniformen Antwortmechanismus und wie erörtert für den Fall der Entstehung fehlender Werte in der Grundgesamtheit, die Varianz unter Aufspaltung der beiden Komponenten, approximiert werden durch:

$$V_S^1 + V_{NR}^1 \approx \frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2 + \frac{1-p_R}{p_R} \cdot \frac{1}{N^2} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2, \quad (3.29)$$

wobei p_R die Antwortwahrscheinlichkeiten darstellen.

Die beiden Komponenten lassen sich auch beschreiben durch:

$$V_S^1 = \frac{1-f}{p_R \cdot f} \cdot \epsilon \quad (3.30)$$

bzw.

$$V_{NR}^1 = \frac{1-p_R}{p_R} \cdot \epsilon \quad (3.31)$$

mit $\epsilon = \frac{1}{N^2} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2$. Die Beiträge der einzelnen Varianzkomponenten sind daher abhängig von den beiden Größen $\eta_S = \frac{1-f}{p_R \cdot f}$ und $\eta_{NR} = \frac{1-p_R}{p_R}$. Bei Annahme der Unverzerrtheit der Schätzung von $\hat{p}_R = r/n$ für p_R und damit unter uniformem Antwortmechanismus und $f = n/N$ können diese Terme umgeformt werden zu:

$$\hat{\eta}_S = \frac{(N-n)}{r} \quad (3.32)$$

bzw.

$$\hat{\eta}_{NR} = \frac{(n-r)}{r} = \frac{m}{r}. \quad (3.33)$$

Es ist zu sehen, dass (3.32) bei kleinem Stichprobenumfang n bzw. großer Grundgesamtheit N groß wird und folglich auch der Term V_S^1 . In diesen Fällen liegt ein kleiner Gesamtauswahlsatz f vor. Die Komponente V_{NR}^1 ist insbesondere abhängig vom Verhältnis fehlender Werte zu den beobachteten Einheiten. Bleibt dieses Verhältnis bei steigendem f (entweder durch Sinken von N oder größer werdendem n) konstant, gewinnt die Komponente V_{NR}^1 einen größeren Anteil, da sie unveränderlich bleibt, die Stichprobenkomponente aber abnimmt. Bei kleinem n jedoch wird die gesamte Varianz hauptsächlich von der Stichprobenkomponente V_S^1 bestimmt. Die Komponente V_{NR}^1 spielt damit nur bei großen Stichprobenumfängen eine Rolle. Außerdem gewinnt sie an Bedeutung bei größer werdendem Verhältnis fehlender Werte zu beobachteter Werte, wobei eine Schätzung gerade dieser Varianzkomponente sehr schwierig wird, wenn das Verhältnis zu Gunsten der fehlenden Werte sehr stark zunimmt. Weiterhin ist zu sehen, dass auch die Stichprobenkomponente V_S^1 von der Anzahl der beobachteten Werte abhängig ist. Dies liegt daran, dass diese Komponente die Variabilität bezüglich der Stichprobenziehung konditioniert auf den Response-Vektor enthält. Nimmt r ab, so unterliegt auch die Berechnung der Komponente V_S^1 einer größeren Anzahl an fehlenden Werten und einer höheren Anzahl an notwendigen Imputationen. Sie nimmt dadurch zu.

Im Falle zufälliger Imputationsverfahren kommt zu den beiden beschriebenen Komponenten noch die Imputationskomponente hinzu. In (3.26) ist dies die dritte Komponente, bezeichnet als \tilde{V}_{Imp}^1 . Nach MASHREGHI et al. (2014, S. 145) entspricht die Ordnung dieser Komponenten die der ersten Komponenten \tilde{V}_S^1 mit $\mathcal{O}(1/n)$. Die Ordnung der zweiten Komponenten \tilde{V}_{NR}^2 bleibt bei $\mathcal{O}(1/N)$. Damit spielen bei einem kleinen Auswahlsatz nur die Komponente \tilde{V}_S^1 und \tilde{V}_{Imp}^1 eine Rolle. Der Term \tilde{V}_{NR}^1 gewinnt weiterhin erst bei einem großen Auswahlsatz an Bedeutung und ist vorher vernachlässigbar (vgl. MASHREGHI et al., 2014, S. 145).

Auch dieser Sachverhalt lässt sich an einem Beispiel verdeutlichen. Bei der (gewichteten) Hot-Deck-Random-Imputation⁷ kommt unter den beschriebenen Annahmen gegenüber (3.29) der Term \tilde{V}_{Imp}^1 hinzu, definiert durch:

$$\tilde{V}_{Imp}^1 = \frac{(1-p_R) \cdot N}{n} \cdot \epsilon, \quad (3.34)$$

(vgl. MASHREGHI et al., 2014, S. 151).

⁷ MASHREGHI et al. (2014) gehen von einer gewichteten Hot-Deck-Random-Imputation aus. Zur Berechnung der Gewichte vgl. MASHREGHI et al., 2014, S. 145.

Unter den bisher getätigten Annahmen kann der Term $\eta_{Imp} = \frac{(1 - p_R) \cdot N}{n}$ umgeformt werden zu:

$$\hat{\eta}_{Imp} = \frac{m \cdot N}{n^2}. \quad (3.35)$$

Die Imputationskomponente der Varianz \tilde{V}_{Imp}^1 nimmt absolut gesehen zu bei einem kleinem Stichprobenumfang und einer großen Anzahl an fehlenden Werten. Hier fallen mehr Imputationen an. Bei großem Stichprobenumfang nimmt sie hingegen deutlich ab. Da sich gegenüber der Mittelwertimputation an den beiden anderen Komponenten nichts geändert hat, gelten hier die zuvor getätigten Erläuterungen. Gegenüber Komponenten \tilde{V}_{NR}^1 wird die Imputationskomponente besonders bei kleinem Auswahlsatz an Bedeutung gewinnen. Um noch detailliertere Informationen über die einzelnen Komponenten zu erhalten, wird im Rahmen der Simulationsstudie eine Monte-Carlo-Varianzzerlegung durchgeführt. Hier sei auf die Ergebnisse in Abschnitt 5.3 verwiesen.

3.5 Varianz bei bestimmten Imputationsmethoden

Im vorherigen Abschnitt wurde schon kurz auf die theoretische Varianz für Mittelwertimputation und Hot-Deck-Random-Imputation eingegangen. In der Literatur finden sich verschiedene Beispiele, wie sich die Varianzen für verschiedene Imputationsmethoden theoretisch aufschreiben lassen. Wie schon der vorherige Abschnitt hat erkennen lassen, sind hierfür jedoch mehr oder weniger starke Annahmen notwendig. Der Grund liegt zum einen darin, dass sich die theoretischen Varianzen ansonsten schwierig oder gar nicht herleiten lassen. Zum anderen sind einige Parameter, z.B. der Antwortmechanismus q , in Praxis oft nicht bekannt. Dieser Abschnitt gibt einen kurzen Überblick über die Varianzen weiterer Imputationsmethoden.

So finden sich in der Literatur in Abhängigkeit der getätigten Annahmen bezüglich Regressionsimputation verschiedene Ausdrucksweisen bei Verwendung des Mittelwertes als interessierende Statistik. Es wird von einer Hilfsvariablen x ausgegangen.

Zum Beispiel geben MASHREGHI et al. (2014) im Rahmen der einfachen Zufallsstichprobe, der Entstehung variabler fehlender Werte in der Grundgesamtheit und uniformem Antwortmechanismus die Varianz für Regressionsimputation an durch:

$$V(\hat{y}_{Reg}) \approx [1 - p_R \cdot f - (1 - p_R) \cdot \text{Cor}^2] \cdot \frac{\sigma^2}{n \cdot p_R}, \quad (3.36)$$

wobei Cor den Korrelationskoeffizient zwischen y und x beschreibt.

KIM und SHAO (2014, S. 60) bzw. KIM (2001) interpretieren den Mittelwertschätzer bei Anwendung von Regressionsimputation gemäß (2.9) als Regressionsschätzer im Sinne von COCHRAN (1977, S. 339). Dementsprechend geben sie die Varianz an durch:

$$V(\hat{y}_{Reg}) = \frac{1}{n} \cdot \sigma^2 + \left(\frac{1}{r} - \frac{1}{n}\right) \cdot \sigma_u^2 = \frac{\sigma^2}{r} \cdot \left(1 - \left(1 - \frac{r}{n}\right) \cdot \text{Cor}^2\right), \quad (3.37)$$

wobei $\sigma_u^2 = (1 - \text{Cor}^2) \cdot \sigma^2$ die Varianz der Störterme ist. Unterschiede zwischen (3.36) und (3.37) können aus verschiedenen Herleitungen und Annahmen resultieren. Statt über

den Weg des Regressionsschätzers verwenden MASHREGHI et al. (2014) die Linearisierung von DEMNATI und RAO (2004) zur Herleitung der Varianz aus (3.36).

Für stochastische Regressionsimputation muss nach KIM und SHAO (2014, S. 60) bzw. KIM (2001) zur Abdeckung der Imputationskomponente (3.37) um den Term $m/n^2 \cdot \sigma_u^2$ erweitert werden:

$$V(\hat{y}_{StochReg}) = \frac{1}{n} \cdot \sigma^2 + \left(\frac{1}{r} - \frac{1}{n}\right) \cdot \sigma_u^2 + \frac{m}{n^2} \cdot \sigma_u^2. \quad (3.38)$$

Auf ähnlichem Weg könnte sich auch die Varianz für den Mittelwertschätzer bei Verhältnisimputation herleiten lassen. Formel (2.7) zeigt, dass hier ein einfacher Verhältnisschätzer vorliegt, für den in der Literatur theoretische Varianzen existieren (vgl. z.B. COCHRAN, 1977, S. 334).

In CHEN und SHAO (2000, S. 121) wird die Varianz bei Nearest-Neighbour-Imputation angegeben durch:

$$V(\hat{y}_{NN}) = \frac{1}{n^2} \cdot E\left[\sum_{i \in R} (1 + d_i)^2 \cdot V(y_i | x_i)\right] + \frac{1}{n^2} V\left[\sum_{i \in R} (1 + d_i) \cdot \xi(x_i)\right], \quad (3.39)$$

wobei d_i die Anzahl der Inanspruchnahme einer Einheit als Spender beschreibt. Sie gehen weiterhin von einer bestimmten unbekanntem und nicht spezifizierten Funktion $\xi(x) = E(y|x)$ aus. Diese kann linear sein, z.B. in der Form $\xi(x) = E(y|x) = \alpha + \beta \cdot x$ oder vollständig unbekannt im Sinne einer nicht parametrischen Variante. Im letzteren Fall wird kein Modell benötigt (vgl. CHEN und SHAO, 2000, S. 116 ff.).

Die Varianz eines Schätzers für den Fall von Fractional-Hot-Deck-Imputation auf Grundlage einer Superpopulation wird in KIM und FULLER (2004) erörtert. Hier erfolgt eine Unterteilung in O Klassen. Grundsätzlich erfolgt die Unterstellung, dass die Beobachtungswerte Realisationen von Zufallsvariablen innerhalb einer Imputationsklasse κ sind, damit gilt $y_i \sim (\mu_\kappa, \sigma_\kappa^2)$. Des Weiteren nehmen sie zusätzlich an, dass diese Realisationen unabhängig von der Stichprobenziehung, wie auch dem Antwortmechanismus, jedoch gegeben der Imputationsklasse betrachtet werden können. Das beschriebene Modell bezieht sich sowohl auf Einheiten mit beobachtetem, wie fehlendem Wert. Damit gilt ebenfalls: $y_i | (s, R) \sim (\mu_\kappa, \sigma_\kappa^2)$. Es wird von noch einigen weiteren Annahmen ausgegangen, wie z.B. dass der Imputationsmechanismus unabhängig von y ist und nur von der Anzahl der Elemente $n_1, n_2 \dots n_O$ und der Anzahl der antwortenden Einheiten $r_1, r_2 \dots r_O$ in den einzelnen Klassen abhängt (vgl. im Detail Theorem 1 in KIM und FULLER, 2004, S. 562). Zudem muss die Wahrscheinlichkeit, dass Element i mindestens einmal als Spender für j verwendet wird, größer als null sein, wobei sich beide Elemente in der gleichen Imputationsklasse befinden. Auch Bedingung (2.15) muss erfüllt sein und w_i muss proportional zu $\{p(i \in S)\}^{-1}$ sein. Dann gilt für die Varianz des Schätzers $\hat{\theta}_{frac}$:

$$V(\hat{\theta}_{frac}) = V\left(\sum_{\kappa=1}^O \sum_{i \in S} w_i \cdot \mu_\kappa\right) + E\left(\sum_{\kappa=1}^O \sum_{i \in R_\kappa} \rho_i^2 \cdot \sigma_\kappa^2\right). \quad (3.40)$$

R_κ ist die Menge der antwortenden Einheiten in Klasse κ und ρ_i definiert nach (2.18).

Der Erwartungswert E bezieht sich auf die gemeinsame Verteilung aus Superpopulationsmodell, Stichprobendesign, Nonresponse- und Imputationsmechanismus. Entsprechendes gilt für die Varianz (vgl. KIM und FULLER, 2004, S. 562).

Die Erläuterungen dieses und des vorherigen Abschnittes zeigen, dass es schwierig ist, im Falle von fehlenden Werten und Imputation die Varianz in verallgemeinerte Form

wie bei vollständigen Beobachtungen aufzuschreiben. Hier ist das Treffen von Annahmen unumgänglich, was die Angabe einer allgemeingültigen theoretischen Varianz für eine bestimmte Imputationsmethode erschwert bzw. unmöglich macht. Dennoch dienen sie zum Teil als Basis für die Herleitung direkter Varianzschätzer. Die Varianz muss geschätzt werden, da die beschriebenen theoretischen Varianzen auf Parameter bezüglich der gesamten Bevölkerung basieren, welche in der Praxis jedoch unbekannt sind. Hier liegt nur die konkrete Stichprobe vor. Die nächsten Abschnitte beschäftigen sich daher mit der konkreten Schätzung der einzelnen Varianzen. Begonnen wird im nächsten Abschnitt mit den direkten Varianzschätzmethoden.

3.6 Direkte Varianzschätzung

Es ist von großem Vorteil, wenn die Varianz eines Schätzers unter Voraussetzung eines bestimmten Stichprobendesigns direkt durch eine Formel wiedergegeben werden kann. Dies ist meist mit geringem computertechnischen Aufwand und Laufzeiten verbunden. Jedoch sind solche Formeln für bestimmte Stichprobendesigns, Schätzstatistiken oder sonstigen gegebenen Bedingungen überhaupt nicht oder nur sehr schwierig herleitbar. Und so verhält es sich auch beim Vorliegen von fehlenden Werten und Imputation.

In Abschnitt 3.4.2 wurde die Zusammensetzung der Varianz eines Schätzers für solch einen Fall erläutert. Eine Möglichkeit wäre die einfache Anwendung des Standardvarianzschätzers $\hat{V}_p(\hat{y})$ aus 3.1 auf den imputierten Datensatz. Dies kann aber zu deutlichen Unterschätzungen der Varianz führen. Der Grund liegt darin, dass dieser sich nur auf die Stichprobenvarianz konzentriert und die Variabilität, welche durch fehlende Werte bzw. Imputation entsteht, nicht erfasst wird (vgl. SHAO und SITTER, 1996, S. 1279, SAIGO et al., 2001, S. 189, RAO und SHAO, 1992, S. 811). Zum anderen hat auch die gewählte Imputationsmethode, insbesondere die deterministischen Verfahren und der Ausfallmechanismus, durchaus negative Effekte auf die Schätzung der Stichprobenvarianz. Damit kann im Normalfall auch eine valide Schätzung der Stichprobenvarianz nicht gewährleistet werden (vgl. RAO und SHAO, 1992, S. 811, LITTLE und RUBIN, 2002, S. 61). Daher sollte auf die Anwendung von Standardvarianzschätzer auf imputierte Daten verzichtet werden. In der Literatur finden sich einige Vorschläge wie eine direkte Varianzschätzung gestaltet werden kann, welche die verschiedenen Komponenten der Varianz berücksichtigt. Häufig werden hierzu, wie bereits im vorherigen Abschnitt beschrieben, bestimmte Annahmen getroffen, welche ihre Anwendung in der Praxis erschweren können. Einige solcher direkten Varianzschätzer werden im Folgenden kurz vorgestellt und einer kritischen Betrachtung unterzogen. Sie hängen insbesondere von der gewählten Imputationsmethode ab. Die direkten Varianzschätzer in diesem Abschnitt werden für den Fall der einfachen Zufallsstichprobe angegeben. Der Grund liegt darin, dass ein großer Anteil dieser Verfahren für dieses Stichprobendesign hergeleitet worden sind. Dies wird entsprechend bei der Beschreibung der Annahmen der einzelnen Verfahren gekennzeichnet. Ihre Übertragbarkeit auf komplexere Stichprobendesigns bzw. Antwortmechanismen ist häufig nicht gegeben oder führt zu nicht zufriedenstellenden Ergebnissen.

3.6.1 Mittelwertimputation

Ein direkter Varianzschätzer für Mittelwertimputation wird in SÄRNDAL und LUNDSTRÖM (2005, S. 171) hergeleitet. Dieser basiert auf der einfachen Zufallsstichprobe. Er leitet diesen Schätzer unter der Annahme der Entstehung fehlender Werte in der Stichprobe ab. Wie in (3.21) kommt es zu einer separaten Ausweisung der Stichprobenkompo-

nente ohne Konditionierung. In diesem Fall gilt nicht die Varianzzerlegung aus Abschnitt 3.4.2. Aber auch hier resultieren zwei Komponenten. Zur Ableitung eines Schätzers für die einzelnen Komponenten greift SÄRNDAL und LUNDSTRÖM (2005) auf die von ihm bezeichnete *Composite-Method* zurück. Für diese beiden Schätzer gilt:

$$\hat{V}_{SAM} = \left(1 - \frac{n}{N}\right) \cdot \frac{s_{yR}^2}{n} \quad (3.41)$$

und

$$\hat{V}_{NR} = \left(\frac{1}{r} - \frac{1}{n}\right) \cdot s_{yR}^2. \quad (3.42)$$

Durch Summierung der beiden Komponenten führt zum Schätzer für die Gesamtvarianz:

$$\hat{V}(\hat{y}_{MW}) = \hat{V}_{SAM} + \hat{V}_{NR} = \left(\frac{1}{r} - \frac{1}{N}\right) \cdot s_{yR}^2, \quad (3.43)$$

mit $s_{yR}^2 = \frac{1}{r-1} \cdot \sum_{i \in R} (y_i - \bar{y}_R)^2$ und \bar{y}_R ist der Mittelwert der beobachteten Einheiten (vgl. SÄRNDAL und LUNDSTRÖM, 2005, S. 171).

Der Varianzschätzer in (3.43) deckt im Gegensatz zum Standard Varianzschätzer die vollständige Varianz ab. Es wird jedoch auch hier die Annahme getroffen, dass der Ausfallmechanismus MCAR ist. Nur in diesem Fall ist nach LITTLE und RUBIN (2002, S. 66 f.) der Schätzer s_{yR}^2 ein unverzerrter Schätzer für die tatsächliche Varianz.

Auch HAZIZA (2010, S. 4) leitet den Varianzschätzer aus (3.43) her. Jedoch nimmt er fehlende Werte in der Grundgesamtheit an und es gilt die Varianzzerlegung aus (3.23). Die einzelnen Komponenten bilden hier die Schätzungen für die Varianzen V_S^1 und V_{NR}^1 aus Abschnitt 3.4.3. Hier gilt:

$$\hat{V}_S^1 = \left(1 - \frac{n}{N}\right) \cdot \frac{s_{yR}^2}{r} \quad (3.44)$$

und

$$\hat{V}_{NR}^1 = \frac{m}{N} \cdot \frac{s_{yR}^2}{r}. \quad (3.45)$$

Die Summe dieser beiden Komponenten ergibt den Varianzschätzer aus (3.43).

3.6.2 Hot-Deck-Random-Imputation

Ein direkter Varianzschätzer für Hot-Deck-Random-Imputation wird in MASHREGHI et al. (2014, S. 164) angegeben:

$$\begin{aligned} \hat{V}(\hat{y}_{HDRI}) &= [1 - \hat{p}_R \cdot f + \hat{p}_R \cdot (1 - \hat{p}_R)] \cdot \frac{s_{yR}^2}{r} \\ &= \left[1 - \frac{r}{N} + \frac{r}{n} \cdot \left(1 - \frac{r}{n}\right)\right] \cdot \frac{s_{yR}^2}{r} \end{aligned} \quad (3.46)$$

mit $\hat{p}_R = r/n$. Auch hier gelten die Annahmen des uniformen Antwortmechanismus.

3.6.3 Verhältnisimputation

Ein Varianzschätzer für Verhältnisimputation wird in RAO (1990) gegeben. Basierend auf dem Konzept der zwei Phasen Stichproben schlägt er den folgenden direkten Varianzschätzer vor:

$$\hat{V}(\hat{y}_{Rat}) = \left(\frac{1}{n} - \frac{1}{N}\right) \cdot s_{yR}^2 + \left(\frac{1}{r} - \frac{1}{n}\right) \cdot s_{eR}^2 \quad (3.47)$$

mit $s_{eR}^2 = \sum_{i \in R} e_i^2 / (r - 1)$, wobei e_i die Imputationsresiduen der antwortenden Einheiten sind, mit $e_i = y_i - \hat{\beta} \cdot x_i$.

Ein weiterer direkter Varianzschätzer beschrieben durch RAO (1990) ist:

$$\begin{aligned} \hat{V}(\hat{y}_{Rat}) &= \left(\frac{1}{n} - \frac{1}{N}\right) \cdot \hat{\beta}^2 \cdot s_{xS}^2 + 2 \cdot \left(\frac{1}{n} - \frac{1}{N}\right) \cdot \hat{\beta} \cdot s_{xeR} \\ &+ \left(\frac{1}{r} - \frac{1}{n}\right) \cdot s_{eR}^2, \end{aligned} \quad (3.48)$$

mit $\hat{\beta} = \sum_{i \in R} y_i / \sum_{i \in R} x_i$ aus Abschnitt 2.2.3, sowie $s_{xeR} = \sum_{i \in R} e_i \cdot x_i / (r - 1)$ und $s_{xS}^2 = \sum_{i \in S} (x_i - \bar{x})^2 / (n - 1)$.

Beide Varianzschätzer sind im Falle großer Stichprobenumfänge unverzerrt. Außerdem basieren sie auf der Annahme, dass der Antwortmechanismus uniform ist. Ist dies nicht der Fall, können starke Verzerrungen resultieren. (vgl. LEE et al., 1994, S. 233).

Ein weiterer direkter Varianzschätzer für Verhältnisimputation wird in SHAO und STEEL (1999) gegeben und gemäß der reversierten Methode hergeleitet. Dieser Schätzer geht von der Annahme aus, dass die fehlenden Werte in der Grundgesamtheit entstehen. Sie nutzen die Tatsache aus, dass sich der innere Vorgang aus (3.23) auf den Stichprobenprozess bezieht. Mittels Taylor Entwicklung und partieller Differentiation leiten sie den folgenden Varianzschätzer für den Totalwert her:

$$\begin{aligned} \hat{V}(\hat{\tau}_{Rat}) &= N^2 \cdot \left(\frac{\bar{x}}{\bar{x}_R}\right)^2 \cdot \left(1 - \frac{r}{N}\right) \cdot \frac{s_e^2}{r} \\ &+ 2 \cdot N^2 \cdot \left(1 - \frac{n}{N}\right) \cdot \left(\frac{\bar{x}}{\bar{x}_R}\right) \cdot \frac{\hat{\beta} \cdot s_{xe}}{n} \\ &+ N^2 \cdot \left(1 - \frac{n}{N}\right) \cdot \frac{\hat{\beta}^2 \cdot s_{xS}^2}{n}, \end{aligned} \quad (3.49)$$

mit $s_e^2 = \frac{1}{r - 1} \cdot (\sum_{i \in R} y_i - \hat{\beta} \cdot x_i)^2$ und $s_{xe} = \frac{1}{r - 1} \cdot [\sum_{i \in R} x_i \cdot (y_i - \hat{\beta} \cdot x_i)]$.

Der Vorteil dieses Schätzers liegt darin, dass er auch bei MAR geeignet ist und sich gut auf geschichtete Designs übertragen lässt. Dies wird noch in der Simulationsstudie zu sehen sein.

Ein letzter Varianzschätzer für Verhältnisimputation wird in SÄRNDAL (1992), LEE et al. (1994) bzw. DEVILLE und SÄRNDAL (1994) beschrieben. Dieser Schätzer ist den sog. modellassistierenden Varianzschätzmethode zuzurechnen. Diese Verfahren basieren auf allgemeinen Regressionsmodellen und deren klassischen Annahmen. Außerdem wurden sie für den Fall der Entstehung fehlender Werte in der Stichprobe hergeleitet. Für den Varianzschätzer für die modellassistierenden Verfahren für den Totalwert gilt:

$$\begin{aligned}\hat{V}(\hat{\tau}_{Rat})_{mod} &= N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \cdot \left[s_{y_S}^2 + C_0 \cdot \hat{\sigma}_e^2 \right] \\ &\quad + N^2 \cdot \left(\frac{1}{r} - \frac{1}{n} \right) \cdot C_1 \cdot \hat{\sigma}_e^2,\end{aligned}\tag{3.50}$$

$$\text{mit } C_0 = \frac{1}{n-1} \cdot \left(\sum_{i \in M} x_i - \frac{\sum_{i \in M} x_i^2}{\sum_{i \in R} x_i} + \frac{1}{n} \cdot \frac{\sum_{i \in M} x_i \cdot \sum_{i \in S} x_i}{\sum_{i \in R} x_i} \right),$$

$$C_1 = \frac{\bar{x}_S \cdot \bar{x}_M}{\bar{x}_R},$$

$$\text{und } \hat{\sigma}_e^2 = \frac{\sum_{i \in R} e_i^2 / (r-1)}{\bar{x}_R \cdot \left[1 - (s_{x_R} / \bar{x}_R)^2 / r \right]}.$$

\bar{x}_M beschreibt dabei den Mittelwert der Hilfsvariable der bezüglich y nicht antwortenden Einheiten definiert durch $\sum_{i \in M} x_i / m$. \bar{x}_R und \bar{x}_S sind die entsprechenden Mittelwerte für die antwortenden Einheiten und die Stichprobe. e_i sind die Residuen mit $e_i = y_i - \hat{\beta} \cdot x_i$. s_{x_R} ist die Standardabweichung der Hilfsvariable der antwortenden Einheiten. $s_{y_S}^2$ ist die Varianz der imputierten Werte der Stichprobe (vgl. LEE et al., 1994, S. 233, SÄRNDAL, 1992, S. 248 ff.).

3.6.4 Regressionsimputation

Ein Schätzer für die Varianz aus (3.36) für den Mittelwertschätzer unter Regressionsimputation wird in MASHREGHI et al. (2014) gegeben. Auch hier wird vom uniformen Antwortmechanismus ausgegangen. Der Schätzer ist definiert als:

$$\begin{aligned}\hat{V}(\hat{y}_{Reg}) &= (1 - \hat{p}_R \cdot f - (1 - \hat{p}_R) \cdot \widehat{\text{Cor}}_R^2) \cdot \frac{s_{y_R}^2}{r} \\ &= \left(1 - \frac{r}{N} - \left(1 - \frac{r}{n} \right) \cdot \widehat{\text{Cor}}_R^2 \right) \cdot \frac{s_{y_R}^2}{r},\end{aligned}\tag{3.51}$$

wobei $\widehat{\text{Cor}}_R^2$ die Korrelation des Untersuchungsmerkmals y und der Hilfsvariable x der antwortenden Einheiten der Stichprobe bezeichnet (vgl. MASHREGHI et al., 2014, S. 160). KIM (2001) nimmt hingegen als Varianzschätzer im Falle der Regressionsimputation die geschätzte Varianz des klassischen Regressionsschätzers, da auch die entsprechende Varianz aus (3.37) zugrunde liegt. Daher gilt für diesen Schätzer:

$$\hat{V}(\hat{y}_{Reg}) = \frac{1}{n \cdot (n-1)} \cdot \sum_{i=1}^n (\hat{y}_i - \bar{y}_I)^2 + \frac{1}{r \cdot (r-2)} \sum_{i \in R} (y_i - \hat{y}_i)^2,\tag{3.52}$$

wobei \hat{y}_i die prognostizierten Werte aus dem Regressionsmodell sind und \bar{y}_I definiert ist durch $\bar{y}_I = 1/n \cdot \sum_{i=1}^n \hat{y}_i$. Der Schätzer aus (3.52) kann nach KIM (2001) umgeformt werden zu

$$\hat{V}(\hat{y}_{Reg}) = \frac{1}{n \cdot (n-1)} \cdot \sum_{i=1}^n (y'_i - \bar{y}_I)^2,\tag{3.53}$$

mit

$$y'_i := \begin{cases} \hat{y}_i, & i \in M, \\ \hat{y}_i + c_r \cdot (y_i - \hat{y}_i), & i \in R, \end{cases} \quad (3.54)$$

sowie

$$c_r = \left[\frac{n \cdot (n-1)}{r \cdot (r-2)} \right]^{1/2}. \quad (3.55)$$

Die Anwendung der Schätzer aus (3.52) bzw. (3.53) erfordert einen starken Regressionszusammenhang. Damit ist zum einen gemeint, dass eine hohe Korrelation zwischen y und x vorhanden sein muss. Außerdem sollte der Ausfallmechanismus MAR sein. Für MCAR kann es sein, dass der angenommene Zusammenhang den Ausfallmechanismus nicht richtig widerspiegelt. Die Anwendbarkeit der beiden Schätzer in (3.51) und (3.52) in Abhängigkeit des Ausfallmechanismus und der Korrelation zwischen y und x werden im Rahmen einer Simulation verglichen. Die Ergebnisse finden sich in Anhang A in Abschnitt A.3.2.

Im Falle von stochastischer Regressionsimputation bei zufälliger Ziehung der Residuen des Regressionsmodells aus den antwortenden Einheiten schlägt KIM (2001) eine Modifikation der beiden Schätzer vor. Es wird von zwei zu schätzenden Koeffizienten α und β ausgegangen. So erfolgt eine Modifikation von (3.52) zu

$$\begin{aligned} \hat{V}(\hat{y}_{StochReg}) &= \frac{1}{n \cdot (n-1)} \cdot \sum_{i=1}^n (\hat{y}_i - \bar{y}_I)^2 \\ &+ \frac{r^{-1} + n^{-2} \cdot m}{r-2} \cdot \sum_{i \in R} (y_i - \hat{y}_i)^2. \end{aligned} \quad (3.56)$$

Bei Varianzschätzer (3.53) wird (3.55) zu

$$c_I = \left[\frac{n \cdot (n-1) \cdot (r^{-1} + n^{-2} \cdot m)}{r-2} \right]^{1/2} \quad (3.57)$$

modifiziert (vgl. KIM, 2001, S. 76 f.).

3.6.5 Nearest-Neighbour-Imputation

Ein Varianzschätzer für die Varianz aus (3.39) im Falle von Nearest-Neighbour-Imputation und eines Mittelwertschätzers wird in CHEN und SHAO (2000) gegeben. Die dort beschriebene Beziehung $\xi = E(y|x)$ wird über ein Modell erfasst. Der Schätzer $\hat{\xi}$ wird entsprechend über dessen Schätzung gewonnen.

Im Falle der einfachen Zufallsstichprobe und einer Imputationsklasse resultiert der folgenden Schätzer:

$$\begin{aligned} \hat{V}(\hat{y}_{NN}) &= \frac{1}{n^2} \sum_{i \in R} (1 + d_i)^2 \cdot [y_i - \hat{\xi}(x_i)]^2 \\ &+ \frac{1}{n \cdot (n-1)} \sum_{i=1}^n \left[\hat{\xi}(x_i) - \frac{1}{n} \cdot \sum_{i=1}^n \hat{\xi}(x_i) \right]^2, \end{aligned} \quad (3.58)$$

(vgl. CHEN und SHAO, 2000, S. 123).

Problematisch ist die Anwendung dieses Schätzer, wenn ein Modell für $\hat{\xi}$ nicht bestimmt werden kann oder falsch spezifiziert ist. Dies kann zu deutlichen Verzerrungen führen. Gerade bei Anwendungen auf realen Daten ist es oft schwierig, diese Beziehung zu schätzen bzw. ein geeignetes Modell aufzustellen. Des Weiteren ist nach BEAUMONT und BOCCI (2009) zu beachten, dass der Varianzschätzer keine Endlichkeitskorrektur berücksichtigt. Daher kann er im Falle von großen Auswahlsätzen beim Ziehen ohne Zurücklegen zu großen Überschätzungen der Varianz führen. BEAUMONT und BOCCI (2009) schlagen daher noch einen weiteren Varianzschätzer im Zusammenhang mit Nereast-Neighbour-Imputation vor, welcher jedoch eher auf die Anwendung von Schätzern für bestimmte Domänen abstellt (vgl. hierzu im Detail BEAUMONT und BOCCI, 2009, S. 403 ff.).

3.6.6 Kritische Bewertung der direkten Varianzschätzer

Auf die Vorteile direkter Varianzschätzer wurde schon eingegangen. Jedoch existieren solche direkte Schätzer nicht für alle Imputationsmethoden. Insbesondere wenn die Imputationsmethode sehr komplex ist oder eine Kombination verschiedener Imputationsmethoden vorliegen, kann eine Herleitung schwierig werden. Aber auch komplexe Stichprobendesigns können für einige dieser Schätzer Probleme bereiten. Dies gilt insbesondere dann, wenn sie auf Basis der einfachen Zufallsstichprobe hergeleitet worden sind, sich aber die übergeordneten Einheiten des Stichprobendesign, z.B. die Schichten, von den Imputationsklassen unterscheiden. In diesem Fall kann keine Aussage getroffen werden, ob ihre Anwendung innerhalb der Schicht oder der Imputationsklasse zu erfolgen hat, um den Anforderungen des Designs oder der Imputation gerecht zu werden. Dieser Sachverhalt gilt besonders dann, wenn die Imputationsklassen zur Gewährleistung des uniformen Antwortmechanismus dienen und dadurch eine Verletzung dieser Annahme auftritt. Dadurch ergibt sich der nächste Nachteil. Die direkten Varianzschätzer basieren häufig auf strengen Annahmen oder auch Modellen. Diese können häufig in der Praxis nicht gegeben sein, wodurch sie nicht mehr oder nur sehr eingeschränkt angewendet werden können. Die Simulationsstudie und Beispiele im Anhang werden dies zeigen. Aufgrund dieser Nachteile sind häufig Alternativen notwendig, welche in den nachfolgenden Kapiteln und Abschnitten beschrieben werden sollen.

Zur Begegnung der Nachteile der direkten und modellassistierenden Varianzschätzer, welche sich auf ihre Anwendung bei komplexen Stichprobendesigns und Imputationsmethoden, der unterschiedlichen Herleitung von Varianzschätzern bezüglich verschiedener Imputationsmethoden und ihrer starken Annahmen konzentrieren, können sog. Resampling-Verfahren eingesetzt werden. Hier existieren eine ganze Reihe unterschiedlicher Methoden, welche im Fokus dieser Arbeit stehen. Aus diesem Grund wird ihnen ein eigenes Kapitel gewidmet, welches nachfolgend behandelt wird.

4 Resampling-Verfahren

4.1 Anwendung der Resampling-Verfahren zur Schätzung der Varianz

Dieses Kapitel beschäftigt sich mit der Anwendung von Resampling-Verfahren zur Schätzung der Varianz bei komplexen Stichprobendesigns, insbesondere im Falle des Auftretens fehlender Werte und Anwendung von Imputation. Einen allgemeinen Überblick dieser Verfahren geben BRUCH (2010), BRUCH et al. (2011), WOLTER (2007), MÜNNICH (2008), SHAO und TU (1995), SHAO (2002), EUROSTAT (2013), DAVISON und SARDY (2004a) sowie DAVISON und SARDY (2007). Dabei wird im nächsten Abschnitt zunächst deren allgemeine Vorgehensweise präsentiert.

4.1.1 Allgemeine Vorgehensweise

Gemeinsam ist allen Resampling-Verfahren, dass aus der vorhandenen Stichprobe Substichproben gezogen werden. Diese werden auch Replikationen genannt. Die verschiedenen Verfahren unterscheiden sich insbesondere dadurch, wie die Substichproben gezogen werden. Für jede dieser Substichproben a wird die interessierende Statistik $\hat{\theta}_a$ berechnet. Dies geschieht auf Basis der für die Substichprobe gezogenen Werte y_i^* , welche durch das für die Resampling-Methode charakterisierende Vorgehen ausgewählt werden. Diese Realisationen werden zusammengefasst zum Vektor y^* . Entsprechend müssen für die meisten Methoden Anpassungen der Designgewichte w_i stattfinden. Dies ist notwendig, da häufig die Substichprobengröße nicht der ursprünglichen Stichprobengröße entspricht oder Endlichkeitskorrekturen für den Fall des Ziehens ohne Zurücklegen einkalkuliert werden müssen. Es resultieren die nach der Resampling-Methode angepassten Designgewichte w_i^* , zusammengefasst durch den Vektor w^* . Damit gilt:

$$\hat{\theta}_a = T(y^*, w^*). \quad (4.1)$$

Ziel dieses grundsätzlichen Vorgehens ist die Verteilung der interessierenden Statistik zu reproduzieren. Um Informationen über die Varianz zu erhalten, wird die Variation dieser Schätzwerte $\hat{\theta}_a$ der einzelnen Substichproben herangezogen. Die Varianz wird demnach geschätzt durch:

$$\hat{V}(\hat{\theta}) = \frac{1}{\Omega} \cdot \sum_{a=1}^{\Omega} (\hat{\theta}_a - \overline{\hat{\theta}_a})^2, \quad (4.2)$$

wobei Ω die Anzahl der gezogenen Substichproben beschreibt.

Dieses beschriebene Vorgehen gilt für den Fall vollständiger Beobachtungen. Treten fehlende Werte auf und findet als deren Kompensation die Imputation statt, müssen diese Vorgänge in die Resampling-Methode integriert werden. Zwei solche Ansätze werden im nächsten Abschnitt vorgestellt.

4.1.2 Resampling-Verfahren bei fehlenden Werten und Imputation

Die entscheidende Frage in diesem Punkt ist, wie die Gesamtvarianz, welche durch das Stichprobendesign, die fehlenden Werte bzw. die Imputation entsteht, durch die Resampling-Verfahren geschätzt werden kann. Des Weiteren ist es wichtig zu wissen, was die Vorteile von Resampling-Verfahren gegenüber den zuvor erläuterten Verfahren sind.

Auch im Zusammenhang mit Resampling-Methoden sind bei der Berücksichtigung von fehlenden Werten einige Annahmen zu berücksichtigen. So gehen viele Arbeiten (vgl. z.B. SHAO und SITTER, 1996, S. 114, RAO und SHAO, 1992, S. 813, SHAO et al., 1998, S. 821), welche im Zusammenhang mit der Anwendung dieser Verfahren stehen, von einem uniformen Antwortmechanismus aus, wie er im Abschnitt 2.1 beschrieben wurde. Zumindest sollten die Antwortwahrscheinlichkeiten innerhalb der gebildeten Imputationsklassen die gleichen sein (vgl. z.B. SHAO und SITTER, 1996, S. 114). Diese Annahme steht häufig im Zusammenhang mit der Gewährleistung der unverzerrten Schätzung des imputierten Punktschätzers und der imputierten Stichprobenvarianz (vgl. RAO und SHAO, 1992, S. 813). Die Simulationsstudie wird jedoch zeigen, dass Resampling-Verfahren bei bestimmten Imputationsmethoden und Verwendung geeigneter Hilfsvariablen deutlich robuster gegenüber der Verletzung der Annahme des uniformen Antwortmechanismus sind, gerade im Vergleich zu den direkten Varianzschätzer.

Eine zweite wichtige Annahme ist das Vorliegen sogenannte Identifikatoren von fehlenden Werten (vgl. SHAO und SITTER, 1996, RAO und SHAO, 1992, SHAO et al., 1998). Die nachfolgenden Ansätze zur Berücksichtigung von fehlenden Werten basieren darauf, dass in der Stichprobe gekennzeichnet ist, ob es sich bei einem Wert um einen tatsächlich beobachteten Wert handelt oder ob dieser Wert imputiert wurde. Dies kann in der Praxis durchaus problematisch sein, wenn Anwendern Datensätze vorliegen, bei welchen diese Unterscheidung nicht mehr ersichtlich ist.

Eine dritte Annahme, welche für die zunächst beschriebenen Resampling-Verfahren getroffen werden soll, ist die Vernachlässigbarkeit der Varianzkomponenten bezüglich fehlender Werte V_{NR}^1 bzw. \tilde{V}_{NR}^1 aus (3.23) respektive (3.26). Wie in Abschnitt 3.4.2 beschrieben, ist dies entweder bei einem geringen Auswahlsatz oder einem fixen Response-Vektor z der Fall. Damit ist zunächst nur die Stichproben- und Imputationskomponente der Varianz von Bedeutung. Bei Anwendung zufälliger Imputationsverfahren werden im Fall großer Auswahlsätze weitere Korrekturen im Zusammenhang mit der Komponenten \tilde{V}_{Imp}^1 notwendig. Dieser Sachverhalt wie auch der Einbezug der Komponenten V_{NR}^1 bzw. \tilde{V}_{NR}^1 werden in Abschnitt 4.7 näher erläutert. Beide Problematiken werden in den folgenden Ausführungen zunächst ausgeklammert. Weiterhin wird wie zuvor unterstellt, dass in den verwendeten Hilfsvariablen keine fehlenden Werte vorliegen.

Nach SHAO (2002) können Resampling-Verfahren, wie in Abschnitt 4.1.1 beschrieben, unter den getroffenen Annahmen mittels zweier Ansätze auf den Fall fehlender Werte und Imputation übertragen werden (vgl. im Folgenden SHAO, 2002, S. 303 ff.). Bei der ersten Möglichkeit spricht SHAO (2002) von *Reimputation*. Hier werden die fehlenden und zuvor imputierten Werte einer Substichprobe $y_i^* \in M^*$ gemäß dem gleichen Verfahren imputiert, welches bei der ursprünglichen Stichprobe angewandt wurde. Dies geschieht auf Basis der beobachteten Werte der Substichprobe $y_i^* \in R^*$ und evtl. vorliegender Hilfsvariablen der Substichprobe. Neben der gleichen Imputationsmethode und den gleichen Hilfsvariablen muss auch berücksichtigt werden, wo die Imputationen der ursprünglichen Stichprobe stattgefunden haben. Wurden diese in Imputationsklassen oder Schichten durchgeführt, so muss dies auch in der Substichprobe geschehen. Die daraus resultierende imputier-

te Variable wird $y^{*,reimp}$ bezeichnet. Hierauf wird wiederum die interessierende Statistik $\hat{\theta}_a^{reimp} = T(y^{*,reimp}, w^*)$ und die Varianz nach (4.2) berechnet.

Wird dies über alle Substichproben hinweg durchgeführt, kann neben der Stichprobenziehung auch der Imputationsprozess reproduziert und damit einhergehende Variation abgedeckt werden. Bei zufälligen Imputationsmethoden können sich jedoch Nachteile ergeben. Werden bei der Reimputation der fehlenden Werte über die Ω Replikationen hinweg unabhängige Zufallsterme benutzt, kann dies zu einer Überschätzung der Varianz führen. Durch die Verwendung unabhängiger Zufallsterme wird nach SHAO (2002) übermäßig viel Variation kreiert.¹ Dies gilt, wenn die Größe der Substichprobe kleiner ist als der ursprüngliche Stichprobenumfang.

Die zweite Möglichkeit, die Varianz durch fehlende Werte bzw. Imputation nach SHAO (2002) zu berücksichtigen, ist die *Adjustierung* der imputierten Werte in jeder Replikation. Hier werden die imputierten Werte $\tilde{y}_{\kappa j}$ der ursprünglichen Stichprobe in jeder Substichprobe gemäß der folgenden Formel angepasst:

$$\tilde{y}_{\kappa j} + E^*(\tilde{y}_{\kappa j}) - E(\tilde{y}_{\kappa j}). \quad (4.3)$$

$E^*(\tilde{y}_{\kappa j})$ beschreibt den Erwartungswert des imputierten Wertes, wenn die Imputation auf Basis der Substichprobe stattfindet. $E(\tilde{y}_{\kappa j})$ ist der Erwartungswert des imputierten Wertes, wenn die Imputation auf Basis der ursprünglichen Stichprobe erfolgt (vgl. SHAO, 2002, S. 307). Der Index κ beschreibt in diesem Abschnitt, wo genau die Imputationen stattfinden. Dies sind häufig aufgrund der schon beschriebenen Annahmen die Imputationsklassen, es können z.B. aber auch die Schichten des Designs sein. In diesem Abschnitt ist dieser Index damit sehr allgemein gehalten. Die auf diesem Wege adjustierte interessierende Variable wird mit $y^{*,adj}$ bezeichnet und es erfolgt die Schätzung von $\hat{\theta}_a^{adj} = T(y^{*,adj}, w^*)$ bzw. der Varianz nach (4.2).

Die Berechnungen von $E^*(\tilde{y}_{\kappa j})$ und $E(\tilde{y}_{\kappa j})$ hängen sehr stark vom gewählten Imputationsverfahren ab. Für deterministische Imputationsverfahren ist $E(\tilde{y}_{\kappa j}) = \tilde{y}_{\kappa j}$ und $E^*(\tilde{y}_{\kappa j}) = \tilde{y}_{\kappa j}^*$, wobei $\tilde{y}_{\kappa j}^*$ der imputierte Wert für den fehlenden Wert ist, wenn die Imputation auf Basis der Substichprobe durchgeführt wird. Da sich in diesem Fall der erste und dritte Term aus (4.3) zu Null subtrahieren, beschreibt (4.3) für deterministische Imputationsverfahren nichts anderes als die Anwendung des Imputationsverfahrens der ursprünglichen Stichprobe auf die Werte der Substichprobe. Dies entspricht dem Vorgehen der Reimputation. Wird dies für jede Substichprobe vorgenommen, kann auch hier der Stichproben- und Imputationsprozess reproduziert werden und die daraus entstehende Varianz abgedeckt werden (vgl. hierzu auch SHAO et al., 1998). Im Falle deterministischer Imputationsverfahren besteht zwischen der Reimputation und der Adjustierung der imputierten Werte kein Unterschied und beide können benutzt werden.

Für zufällige Imputationsverfahren ergeben sich jedoch Unterschiede zwischen den Verfahren. Für eine einfache Hot-Deck-Random-Imputation stellt $E(\tilde{y}_{\kappa j})$ den Mittelwert der beobachtbaren Werte der ursprünglichen Stichprobe in der jeweiligen Kategorie κ dar. $E^*(\tilde{y}_{\kappa j})$ ist der Mittelwert der beobachtbaren Werte der Substichprobe in der jeweiligen Kategorie κ . Gleiches gilt für Fractional-Imputation, falls die dem Spender einer Kategorie zugeordneten Gewichte gleich groß sind. Falls dies nicht der Fall ist, wird die Berechnung der beiden Größen deutlich komplexer und hängt im wesentlichen davon ab wie die Gewichte bzw. Fractions genau berechnet werden. Diese Vorgehensweise muss bei

¹ Die zweite von SHAO (2002) vorgeschlagene Methode zur Reimputation, die Verwendung gleicher Zufallsterme in der ursprünglichen Imputation, wird bei den Imputationsmethoden dieser Arbeit nicht berücksichtigt.

Imputationsmethode	$E^*(\tilde{y}_{\kappa j})$	$E(\tilde{y}_{\kappa j})$
Deterministisch	$\tilde{y}_{\kappa j}^*$	$\tilde{y}_{\kappa j}$
HDMI	$\bar{y}_{R_\kappa}^*$	\bar{y}_{R_κ}
Stochastische Regressionsimputation	vgl. (4.5)	vgl. (4.4)
Fractional Imputation	$\bar{y}_{R_\kappa}^*$	\bar{y}_{R_κ}

Tabelle 4.1: Adjustierung imputierter Werte

der Berechnung des jeweiligen Mittelwertes berücksichtigt werden. Für stochastische Regressionsimputation setzen sich $E(\tilde{y}_{\kappa j})$ und $E^*(\tilde{y}_{\kappa j})$ aus zufälligen und deterministischen Komponenten zusammen. Es ergibt sich:

$$E(\tilde{y}_{\kappa j}) = \tilde{y}_{\kappa j, reg} + 1/r_\kappa \cdot \sum_{i \in R_\kappa} (y_{\kappa i} - \tilde{y}_{\kappa i, reg}) \quad (4.4)$$

und

$$E^*(\tilde{y}_{\kappa j}) = \tilde{y}_{\kappa j, reg}^* + 1/r_\kappa^* \cdot \sum_{i \in R_\kappa} (y_{\kappa i}^* - \tilde{y}_{\kappa i, reg}^*), \quad (4.5)$$

wobei $\tilde{y}_{\kappa j, reg}$ den imputierten Wert für die ursprüngliche Stichprobe darstellt, wenn Regressionsimputation verwendet wurde und $\tilde{y}_{\kappa j, reg}^*$ den jeweilige Wert für die Substichprobe. r_κ ist die Anzahl der beobachteten Einheiten in Kategorie κ und $1/r_\kappa^*$ die Anzahl beobachtbarer Einheiten in Kategorie κ in der Substichprobe. $1/r_\kappa \cdot \sum_{i \in R_\kappa} (y_{\kappa i} - \tilde{y}_{\kappa i})$ beschreibt den Mittelwert der Residuen der beobachtbaren Werte wenn Regressionsimputation durchgeführt wurde und $1/r_\kappa^* \cdot \sum_{i \in R_\kappa} (y_{\kappa i}^* - \tilde{y}_{\kappa i}^*)$ den entsprechenden Wert für die Substichprobe.

Sie decken die zufällige Komponente ab. Eine Übersicht über die jeweiligen Werte für $E^*(\tilde{y}_{\kappa j})$ und $E(\tilde{y}_{\kappa j})$ mit Bezug auf die in dieser Arbeit verwendeten Imputationsmethoden wird in Tabelle 4.1 gegeben.

Im Falle zufälliger Imputationsverfahren führen die Adjustierung der imputierten Werte und die Reimputation zu unterschiedlichen Ergebnissen. Bei der Reimputation werden fehlende Werte der Substichprobe zufällig mit beobachtbaren Werten der Substichprobe bzw. deren Komponenten, wie Residuen, imputiert. Diese unabhängige Verwendung von Zufallstermen über die Ω Replikationen hinweg liefert nach SHAO (2002), wie erläutert, bei Substichprobengrößen kleiner als die ursprüngliche Stichprobengröße Überschätzungen der Varianz. Bei der Adjustierung wird diesem Umstand Rechnung getragen, indem imputierte Werte in Richtung des Erwartungswertes $E^*(\tilde{y}_{\kappa j})$ angepasst werden. Dies ist so zu verstehen, dass ein imputierter Wert in der Substichprobe erhöht bzw. verringert wird, wenn dessen Erwartungswert in der Substichprobe größer bzw. kleiner ist als der analoge Wert in der ursprünglichen Stichprobe.

Der imputierte Wert $\tilde{y}_{\kappa j}$ ist über die Ω Replikationen gegeben, falls der entsprechende fehlende Wert in die Substichprobe gelangt. Auch der Erwartungswert $E(\tilde{y}_{\kappa j})$ ist über die Replikationen unveränderlich. Die einzige Größe, die damit umfassenderen Schwankungen unterliegt, ist der Erwartungswert $E^*(\tilde{y}_{\kappa j})$. Der zufällige Anteil dieser Größe ist jedoch für fehlende Einheiten einer Replikation in einer bestimmten Kategorie κ durch die Mittelwertbildung immer der Gleiche. Durch die Anpassung der imputierten Werte in Richtung dieses Erwartungswertes der Substichprobe entsteht weniger Variation. Dies gilt im Vergleich dazu, dass bezüglich verschiedener fehlender Werte sowie für gleiche fehlende Werte

über die Replikationen hinweg, mit wechselnden zufälligen Termen reimputiert wird. Die Varianz wird dadurch nicht überschätzt.

Die Reimputation besitzt hingegen Vorteile bezüglich ihrer einfacheren Implementierbarkeit. Das originäre Imputationsschema der ursprünglichen Stichprobe kann direkt auf die Daten der Substichprobe angewendet werden. Die beiden Erwartungswerte können unter Umständen schwierig oder gar nicht zu bestimmen sein. Zudem kann im Falle einer zu imputierenden kategorialen Variable Formel (4.3) nicht angewendet werden.

Ein in der Praxis sehr interessanter Bereich der Anwendung von Resampling-Verfahren ist die Kombination von Imputationsverfahren. Damit ist gemeint, dass in unterschiedlichen Imputationsklassen verschiedene Imputationsmethoden angewendet werden. Für Resampling-Verfahren bedeutet dies, dass auch die zum Imputationsverfahren gehörende Reimputation bzw. Adjustierung der imputierten Werte in der entsprechenden Imputationsklasse in der Substichprobe angewendet wird. Wird z.B. in bestimmten Imputationsklassen die Mittelwertimputation vorgenommen, kann in diesen Klassen eine Reimputation mittels dieser Methode stattfinden. Wird in den anderen Klassen hingegen Hot-Deck-Random-Imputation durchgeführt, so muss in diesen Klassen eine Adjustierung der imputierten Werte gemäß (4.3) mit den entsprechenden Erwartungswerten aus Tabelle 4.1 vollzogen werden. So kommt es zu einer kombinierten Anwendung von Reimputation und Adjustierung.

Die bis jetzt beschriebenen Verfahren gehen von einem einstufigen Design aus. Im Falle zusätzlicher Stufen gilt es, diese Ansätze entsprechend anzupassen.

4.1.3 Resampling-Verfahren bei mehrstufigen Stichprobendesigns

Wie in Abschnitt 3.1 beschrieben, kann es für Varianzschätzer unabdingbar sein, alle Stufen eines mehrstufigen Stichprobendesigns zu berücksichtigen. Der zentrale Punkt in diesem Zusammenhang ist ein großer Varianzanteil auf der zweiten bzw. noch tiefer liegender Stufen des Stichprobendesigns. In solchen Fällen müssen die Varianzschätzer die entstehende Varianz aller Stufen abdecken.

Dies gilt auch für Resampling-Methoden. Eine Möglichkeit zur Umsetzung dieser Verfahren und einer Gewährleistung eines Varianzschätzers, welcher alle Stufen eines Stichprobendesigns abdeckt, ist es die Resampling-Methode separat auf allen Stufen anzuwenden (vgl. im Folgenden auch BRUCH, 2010 oder BRUCH et al., 2011). Im Falle eines zweistufigen Stichprobendesigns ohne Schichtung und einer linearen Statistik θ gilt:

$$\begin{aligned} \hat{V}_{mult}(\hat{\theta}_{2St}) &= (1 - f_1) \cdot \sum_{a=1}^{\Omega_1} \left(\hat{\theta}_a - \frac{1}{\Omega_1} \cdot \sum_{v=1}^{\Omega_1} \hat{\theta}_v \right)^2 \\ &\quad + \frac{L}{l} \cdot \sum_{a=1}^{\Omega_1} (1 - f_{2,d}) \cdot \sum_{o=1}^{\Omega_2} \left(\hat{\theta}_{d,o} - \frac{1}{\Omega_2} \sum_{j=1}^{\Omega_2} \hat{\theta}_{d,j} \right)^2, \end{aligned} \quad (4.6)$$

mit $f_1 = l/L$, $f_{2,d} = n_d/N_d$ und $\hat{\theta}_a = T(y_1^*, w_1^*)$, mit $w_1^* = w_d^* \cdot w_{di}$. y_1^* stellt die Beobachtungen der interessierenden Variablen für Elemente der letzten Stufe dar, welche aber über das Resampling von Elementen der ersten Stufe gewonnen wurden. w_d^* sind die nach der Resampling-Methode angepassten Designgewichte der ersten Stufe. w_{di} sind die nicht angepassten Gewichte der zweiten Stufe, da auf der ersten Stufe alle Einheiten einer gezogenen PSU berücksichtigt werden. Weiterhin gilt $\hat{\theta}_{d,o} = T(y_2^*, w_{di}^*)$. y_2^* sind

die Realisation von Elementen der letzten Stufe, welche auch über das Resampling von Elementen der letzten Stufe erhalten wurden. w_{di}^* sind die nach der Resampling-Methode angepassten Designgewichte der zweiten Stufe. Ω_1 beschreibt die Anzahl der Replikationen auf der ersten und Ω_2 die Anzahl der Replikationen auf der zweiten Stufe. Es ist denkbar, Ω_2 zwischen den PSU zu variieren. Zur besseren Vergleichbarkeit wird hier auf eine gleichbleibende Anzahl an Replikationen über die PSU hinweg abgestellt.

Die Anwendung von Formel (4.6) setzt voraus, dass Resampling-Verfahren z.B. sinnvoll innerhalb der PSU angewendet werden können. Dies muss aber nicht unbedingt so sein. Vielmehr kann es notwendig werden, den Punktschätzer der Resampling-Methode auf der zweiten Stufe $\hat{\theta}_o$ über die PSU hinweg zu berechnen, wobei die Ziehung der Substichproben weiterhin innerhalb der PSU erfolgt. Dieses Vorgehen setzt voraus, dass die Endlichkeitskorrekturen $1 - f_1$ und $1 - f_{2,d}$ in die Berechnung der Punktschätzer der Resampling-Methode integriert werden können. Dies gilt insbesondere für die Endlichkeitskorrektur der zweiten Stufe.² Im einfachsten Fall konstanter Auswahlätze innerhalb der PSU kann sie über ihre Wurzel in die Gewichtungsberechnung miteinkalkuliert werden (vgl. hierzu auch die Erläuterungen in RAO und SHAO, 1996, MCCARTHY, 1966 und VALLIANT, 1987). Bei unterschiedlichen Auswahlätzen werden die Gewichtsadjustierungen aufwendiger, wie es z.B. in Abschnitt 4.6.3 beschrieben wird.

Gemäß dieses Vorgehens modifiziert sich Formel (4.6) zu:

$$\begin{aligned} \hat{V}_{mult}(\hat{\theta}_{2St}) &= \sum_{a=1}^{\Omega_1} \left(\hat{\theta}_a - \frac{1}{\Omega_1} \cdot \sum_{v=1}^{\Omega_1} \hat{\theta}_v \right)^2 \\ &+ \frac{L}{l} \cdot \sum_{o=1}^{\Omega_2} \left(\hat{\theta}_o - \frac{1}{\Omega_2} \sum_{j=1}^{\Omega_2} \hat{\theta}_j \right)^2. \end{aligned} \quad (4.7)$$

Liegt auf der ersten Stufe eine Schichtung vor, kann der Term L_h/l_h nur im Falle konstanter Auswahlätze der PSU über die Schichten hinweg, z.B. im Falle einer proportionalen Allokation, vor die Summe gezogen werden. Ansonsten ist dieses Vorgehen nicht anwendbar.

Die bisherigen Ausführungen zu den Resampling-Methoden bei mehrstufigen Stichprobendesigns gelten ausschließlich für vollständige Beobachtungen. Bei Auftreten fehlender Werte muss wieder auf die im vorherigen Abschnitt beschriebenen Verfahren der Reimputation und Adjustierung der imputierten Werte zurückgegriffen werden. Diese beiden Methoden müssen entsprechend der Formeln (4.6) und (4.7) innerhalb der Resampling-Methode auf allen Stufen des Stichprobendesigns durchgeführt werden. Wird ein kleiner Auswahlatz und ein deterministisches Imputationsverfahren angenommen, so ist in den folgenden Erläuterungen nur die Stichprobenkomponente V_S^1 von Bedeutung. In diesem Zusammenhang ist entscheidend, dass selbst bei ausschließlicher Vornahme der Imputation auf der letzten Stufe des Stichprobendesigns auch die Varianzen übergeordneter Stufen beeinflusst werden. So werden z.B. im zweistufigen Design³ bei der Berechnung von Totalwerten die imputierten Einheiten der letzten Stufen zu Gesamtwerten der einzelnen

² Für die Endlichkeitskorrektur der ersten Stufe gilt diese Aussage z.B. im Falle des Vorliegens einer Schichtung auf der ersten Stufe. Da in (4.7) keine Schichtung vorliegt, könnte die Varianz der ersten Stufe wie in (4.6) angegeben werden. Aus Gründen der Einheitlichkeit und da der Schätzer (4.7) im späteren Verlauf im Rahmen einer Schichtung auf der ersten Stufe relevant ist, wird hier auf eine Einberechnung in den Punktschätzer abgestellt.

³ vgl. hierzu auch die Ausführungen zum mehrstufigen Stichprobendesign bei vollständigen Beobachtungen in Abschnitt 3.1.

PSU aufsummiert. Durch diesen Vorgang ist die Varianz dieser Stufe von den Imputationen betroffen, sogar dann, wenn auf der Stufe selbst keine Imputationen vorgenommen worden sind. Daher müssen bei separater Anwendung von Resampling-Verfahren auf den einzelnen Stufen Reimputationen und Adjustierungen der imputierten Werte nicht nur auf der zweiten Stufe, sondern auch auf der ersten Stufe durchgeführt werden. Nur auf diesem Wege ist es möglich, die Varianz korrekt abzudecken.

Aus diesem Grund müssen $\hat{\theta}_a$ und $\hat{\theta}_{d,o}$ bzw. $\hat{\theta}_o$ in (4.6) und (4.7) durch $\hat{\theta}_a^{reimp}$, $\hat{\theta}_{d,o}^{reimp}$, $\hat{\theta}_o^{reimp}$ bzw. $\hat{\theta}_a^{adj}$, $\hat{\theta}_{d,o}^{adj}$, $\hat{\theta}_o^{adj}$ ersetzt werden. Diese beschreiben die korrespondierenden reimputierten bzw. adjustierten Schätzer.

Nachteilig kann sich bei diesen Verfahren der größere Aufwand auswirken, der sowohl hinsichtlich der Programmierung, wie auch durch längere Laufzeiten entsteht. Der Hauptgrund liegt darin, dass die einzelnen Resampling-Verfahren auf allen Stufen des Stichprobendesigns nach Formel (4.6) bzw. (4.7) separat angewendet werden müssen. Dies gilt ebenfalls für die Verfahren zur Berücksichtigung der Imputation aus dem vorherigen Abschnitt. Im Rahmen dieser Arbeit wird lediglich die Methode des Balanced-Repeated-Replication, welche in Abschnitt 4.6 erläutert wird, nach diesem Prinzip bei mehrstufigen Stichprobendesigns angewendet.

Zudem wird noch eine weitere Resampling-Methode im Falle eines mehrstufigen Designs berücksichtigt, welche alle Stufen des Stichprobendesigns innerhalb der Berechnung des Punktschätzers integriert. Dadurch muss das Verfahren nicht auf allen Stufen des Stichprobendesigns separat angewendet werden. Hierbei handelt es sich um den Rescaling-Bootstrap ohne Zurücklegen, welcher in Abschnitt 4.5 dargestellt wird. Weitere in dieser Arbeit vorgestellte Resampling-Verfahren werden im Falle mehrstufiger Designs nicht berücksichtigt. Der Grund liegt in ihren erzielten Resultaten für das einstufige Stichprobendesign im Rahmen der Simulationsstudie in Abschnitt 5.2, aber auch in Voruntersuchungen bezüglich ihrer Anwendung bei mehrstufigen Stichprobendesigns.

4.1.4 Vor- und Nachteile von Resampling-Verfahren

Die großen Vorteile von Resampling-Verfahren gegenüber den zuvor erläuterten Varianzschätzmethoden werden in SHAO (2002, S. 303) erläutert. So ist ein zentraler Vorteil, dass im Vergleich zu den direkten oder modellassistierenden Methoden nicht für jedes Problem eine neue theoretische Formel hergeleitet werden muss. Das grundlegende Vorgehen wird beibehalten. Für Änderungen der interessierenden Statistik oder der Imputationsmethode sind häufig nur geringe Anpassungen notwendig. Wie zuvor erläutert, ergeben sich dadurch bei den Imputationsmethoden große Vorteile. Durch ihr unkompliziertes Vorgehen kann auch die Behandlung von komplexen Problemen, wie z.B. das Vorliegen mehrstufiger Stichprobendesigns oder die Kombination von Imputationsverfahren auf einfachem Wege gelöst werden. Gegenüber den modellassistierenden Methoden sind die Verfahren auch bei Abweichung vom Modell robust. Des Weiteren sind die getroffenen Annahmen bei Auftreten fehlender Werte meist nicht so streng wie im Falle der direkten Varianzschätzung.

Dem gegenüber stehen insbesondere der computertechnische Aufwand, insbesondere in Form längerer Rechenzeiten (vgl. z.B. MÜNNICH, 2008, S. 327). Diese hängen von der Anzahl der Replikationen ab. Häufig ist ein Konflikt zwischen den längeren Berechnungszeiten und der Genauigkeit der Messung gegeben. Diese steigt mit der Zunahme an Replikationen. Aufgrund immer leistungsstärkerer Computer muss dieser Kritikpunkt zunehmend relativiert werden. Aber auch ihre Implementierung ist aufwendiger im Vergleich zu direkten Varianzschätzern oder modellassistierenden Methoden. Diese lassen sich zu-

mindest für einfache Imputationsmethoden häufig durch eine einfache Formel ausdrücken. Für Resampling-Verfahren ist der Programmieraufwand deutlich höher. In den nachfolgenden Abschnitten werden verschiedene Resampling-Verfahren vorgestellt. Die erste vorzustellende Methode wird als Random-Groups bezeichnet.

4.2 Random-Groups

Dieses Verfahren wird insbesondere in WOLTER (2007) erläutert und stellt die einfachste Form eines Resampling-Verfahrens dar. Dabei ergibt sich das folgende Vorgehen (vgl. im Folgenden SHAO und TANG, 2011): Es werden zufällig G Gruppen von ungefähr gleicher Größe gebildet. Jede dieser Gruppen bildet eine Substichprobe mit dem dazugehörigen Vektor der interessierenden Variable y_g . Es ergeben sich dadurch G Replikationen. Die Designgewichte $w_{i,g}$, der sich in einer Gruppe bzw. Replikation befindlichen Einheiten, werden bestimmt durch $w_{i,g} = w_i \cdot G$, zusammengefasst zum Gewichtevektor w_g . Anschließend wird der Punktschätzer jeder Substichprobe durch $\hat{\theta}_g = T(y_g, w_g)$ berechnet. Für den Varianzschätzer gilt:

$$\hat{V}_{RG}(\hat{\tau}) = \frac{1}{G \cdot (G - 1)} \cdot \sum_{g=1}^G \left(\hat{\theta}_g - \frac{1}{G} \cdot \sum_{l=1}^G \hat{\theta}_l \right)^2, \quad (4.8)$$

(vgl. SHAO und TANG, 2011, S. 509).

Nach WOLTER (2007) sowie SHAO und TANG (2011) muss das verwendete Stichprobendesign innerhalb der Gruppen widergespiegelt werden. Damit ist zum Beispiel gemeint, dass im Falle der geschichteten Zufallsstichprobe eine Gruppe alle H Schichten beinhalten sollte. Die Gruppen werden in diesem Fall über die Schichten hinweg gebildet. Dieses Vorgehen macht den Einbezug von Endlichkeitskorrekturen zur Berücksichtigung des Ziehens ohne Zurücklegen schwierig. Dies gilt insbesondere wenn die Auswahlätze z.B. zwischen den Schichten variieren. Der in SHAO und TANG (2011) dargestellte Schätzer geht daher von kleinen Auswahlätzen bzw. vernachlässigbaren Endlichkeitskorrekturen aus.

Zur Berücksichtigung fehlender Werte und Imputation stellen SHAO und TANG (2011) insbesondere auf die Reimputation ab. Bei Anwendung dieses Ansatzes auf den Vektor y_g , unter Rückgriff evtl. vorliegender Hilfsvariablen oder Imputationsklassen, resultiert für jede Gruppe und damit jede Substichprobe der reimputierte Vektor $y_{g,rimp}$. Für den Punktschätzer gilt:

$$\hat{\theta}_g = T(y_{g,rimp}, w_g). \quad (4.9)$$

Die Varianz wird weiterhin nach (4.8) berechnet. Im Falle zufälliger Imputationsverfahren sollte jedoch auf die Adjustierung der imputierten Werte abgestellt werden. Dadurch ergibt sich der adjustierte Vektor $y_{g,adjust}$. Der Punktschätzer wird berechnet durch:

$$\hat{\theta}_g = T(y_{g,adjust}, w_g). \quad (4.10)$$

Die Methode der Random-Groups zeigt gegenüber den anderen Resampling-Methoden einen großen Nachteil. Da für jede Substichprobe immer nur eine Gruppe verwendet wird, können die Schätzungen in jeder Replikationen nur auf sehr wenigen Einheiten basieren, welche entsprechend hochgewichtet werden. Dies gilt insbesondere, wenn viele Gruppen

mit wenigen Elementen gebildet werden. Varianzschätzungen können hier sehr instabil sein, da die geringe Anzahl an Elementen keine valide Grundlage mehr bildet. Treten zusätzlich noch fehlende Werte auf, welche zu imputieren sind, kann die Schätzung sehr ungenau werden. Daher wären eher größere wenige Gruppen sinnvoll. Jedoch gehen wenige Gruppen mit einer geringen Anzahl von Substichproben einher, worunter auch wieder die Qualität der Schätzung leiden kann. Um diesem Nachteil zu begegnen und stabilere Gruppen zu erhalten, gehen die Resampling-Methoden, welche im nächsten Abschnitt behandelt werden, bei der Substichprobenbildung von einer Löschung einzelner Einheiten bzw. Gruppen aus. Diese Methoden werden unter dem Begriff Jackknife subsumiert.

4.3 Jackknife

Grundlegend ist der Methode des Jackknife, dass eine Substichprobe dadurch gebildet wird, indem Elemente der ursprünglichen Stichprobe für die Replikation gelöscht werden. So wird beim Delete-1-Jackknife im Falle einer einfachen Zufallsstichprobe immer ein Element der ursprünglichen Stichprobe gelöscht, bis alle Elemente einmal ausgelassen wurden. Dadurch ergeben sich n Replikationen. Der Vektor der interessierenden Variable nach Löschung eines Elementes wird als y_{-1} bezeichnet. Es wird hier davon ausgegangen, dass die Elemente direkt gelöscht werden und nicht mit einer Gewichtszuweisung von Null gearbeitet wird.

Zur Kalkulation der Statistik in jeder Substichprobe muss der Tatsache Rechnung getragen werden, dass sie nicht auf Basis der vollständigen Stichprobengröße n sondern auf $n - 1$ Einheiten geschätzt wird. Dies geschieht über die Adjustierung der Designgewichte. Demnach werden die Gewichte der nicht gelöschten Einheiten angepasst durch $w_{i,-1} = w_i \cdot n / (n - 1)$. Die Gewichte werden durch den Gewichtevektor w_{-1} zusammengefasst.

Die interessierende Statistik wird anschließend berechnet durch $\hat{\theta}_{-i} = T(y_{-1}, w_{-1})$. $\hat{\theta}_{-i}$ stellt damit die interessierende Statistik dar, welche auf Basis der Replikation unter Auslassung von Element i berechnet wurde. Die Varianz wird anschließend berechnet durch:

$$\hat{V}_{d1JK}(\hat{\theta}) = \frac{n-1}{n} \cdot \sum_{i=1}^n \left(\hat{\theta}_{-i} - \frac{1}{n} \cdot \sum_{j=1}^n \hat{\theta}_{-j} \right)^2, \quad (4.11)$$

(vgl. SHAO und TU, 1995, S. 4 ff.).

Auch auf stratifizierte Stichprobendesigns kann der Delete-1-Jackknife einfach angewendet werden. Hier existieren zwei Möglichkeiten. So kann der Delete-1-Jackknife-Punktschätzer über die Schichten hinweg berechnet werden. In diesem Fall wird bei Löschung des Elementes i in Schicht h die Gewichte des Elementes j in Schicht t wie folgt adjustiert:

$$w_{t,j}^{(hi)} := \begin{cases} w_{tj}, & \text{wenn } t \neq h, \\ w_{hj} \cdot n_h / (n_h - 1), & \text{wenn } t = h, j \neq i, \end{cases} \quad (4.12)$$

(vgl. DAVISON und SARDY, 2004a, S. 14). Problematisch stellt sich bei diesem Vorgehen die Integration der schichtspezifischen Endlichkeitskorrektur $1 - f_h$ dar. Im Falle eines konstanten Auswahlsatzes über Schichten hinweg könnte diese Korrektur über ihre Wurzel in die Punktschätzung integriert werden. Andernfalls können beim Modell ohne Zurücklegen Überschätzungen auftreten. Für lineare Statistiken ist es auch möglich, den

Delete-1-Jackknife innerhalb der einzelnen Schicht anzuwenden, d.h. die Elemente werden nacheinander in jeder Schicht gelöscht und die einzelnen Punktschätzer innerhalb der Schichten berechnet. Dann wird der Delete-1-Jackknife-Varianzschätzer wie folgt berechnet:

$$\hat{V}_{d1JK, strat}(\hat{\theta}) = \sum_{h=1}^H \frac{(1 - f_h) \cdot (n_h - 1)}{n_h} \sum_{i=1}^{n_h} \left(\hat{\theta}_{h,-i} - \overline{\hat{\theta}}_h \right)^2. \quad (4.13)$$

$\hat{\theta}_{h,-i}$ entspricht in diesem Fall der interessierenden Größe in Schicht h unter Auslassung des Elementes i . $\overline{\hat{\theta}}_h$ ist der Mittelwert dieser Statistik über die verschiedenen Replikationen in Schicht h und f_h wie zuvor der Auswahlatz in der entsprechenden Schicht h (vgl. DAVISON und SARDY, 2004a, S. 14, MÜNNICH, 2008, S. 326). Ein Vorteil dieses Vorgehens ist die einfache Berücksichtigung der Endlichkeitskorrektur $1 - f_h$, wobei hier der Auswahlatz zwischen den Schichten variieren kann. In diesem Fall sind nur die Gewichte der Einheiten der betrachteten Schicht h von Interesse. Diese werden wie in (4.12) angepasst durch:

$$w_{hj,-1} = w_{hj} \cdot n_h / (n_h - 1). \quad (4.14)$$

Die einzelnen Gewichte werden wieder zum Vektor $w_{h,-1}$ zusammengefasst. Für $\hat{\theta}_{h,-i}$ gilt $\hat{\theta}_{h,-i} = T(y_{h,-1}, w_{h,-1})$, wobei $y_{h,-1}$ der Vektor der interessierenden Variable in Schicht h unter Auslassung eines Elementes darstellt.

Der Delete-1-Jackknife besitzt grundsätzlich zwei größere Nachteile. Zum einen setzt er glatte Statistiken voraus (vgl. MÜNNICH, 2008, S. 326). Im Falle nicht linearer Statistiken können Inkonsistenzen resultieren (vgl. SHAO und WU, 1989, S. 1176). Das bekannteste Beispiel hierfür ist der Median und wird in EFRON (1982) näher erläutert.

Zudem kann bei großen Stichprobengrößen n der Rechenaufwand erheblich ansteigen, da hier eine große Anzahl an Replikationen resultiert. Dies kann ihn insbesondere für Erhebungen in der Amtlichen Statistik unpraktikabel machen. Auch im Rahmen der Simulationsstudie dieser Arbeit kann der Delete-1-Jackknife aufgrund dieses hohen Rechenaufwandes nicht berücksichtigt werden.

4.3.1 Delete-a-Group-Jackknife

Eine Möglichkeit den Aufwand von n Replikationen zu reduzieren, ist, in einer Substichprobe mehr als ein Element zu löschen. Auf diesem Wege sind weniger Replikationen notwendig um alle Elemente einmal auszulassen. Ein solches Vorgehen wird z.B. mit dem Delete-a-Group-Jackknife erreicht. Diese Modifikation des Delete-1-Jackknife wird z.B. in KOTT (1998) oder KOTT (2001) beschrieben. Statt eines Elementes werden hier Gruppen von Elementen in einer Substichprobe gelöscht. Die verbliebenen Werte der interessierenden Variable werden als y_{-g} bezeichnet. Zur Gruppenbildung wird in dieser Arbeit eine zufällige Einteilung durchgeführt. Wichtig ist, dass die Gruppen ungefähr die gleiche Größe besitzen und ein Element nicht in mehreren Gruppen auftaucht (vgl. BIENAS et al., 2003, S. 539).

Wie zuvor beim Delete-1-Jackknife müssen auch hier die Designgewichte der nicht gelöschten Einheiten zur Berechnung des Delete-a-Group-Jackknife-Punktschätzers angepasst werden. Hier muss jedoch der Tatsache Rechnung getragen werden, dass die Löschung einer ganzen Gruppe stattfindet. Es gilt:

$$w_{j,-g} = w_j \cdot n / (n - n_g), \quad (4.15)$$

wobei n_g die Anzahl der Elemente ist, welche sich in Gruppe g befinden (vgl. KOTT, 2001, S. 522). Die Gewichte werden zum Gewichtevektor w_{-g} zusammengefasst. Wiederum wird anschließend die interessierende Statistik berechnet durch $\hat{\theta}_{-g} = T(y_{-g}, w_{-g})$. $\hat{\theta}_{-g}$ ist damit die interessierende Statistik, welche ohne die Elemente der Gruppe g berechnet wurden. Für die geschätzte Varianz mittels des Delete-a-Group-Jackknife gilt:

$$\hat{V}_{dgJK}(\hat{\tau}) = \frac{G-1}{G} \cdot \sum_{g=1}^G \left(\hat{\theta}_{-g} - \frac{1}{G} \cdot \sum_{l=1}^G \hat{\theta}_{-l} \right)^2, \quad (4.16)$$

wobei G die Anzahl der gebildeten Gruppen (vgl. SHAO und TU, 1995, S. 195) ist. Bei geschichteten Designs kann entsprechend der beiden Möglichkeiten des Delete-1-Jackknife vorgegangen werden, wobei einzelne Gruppen in den Schichten gelöscht werden und eine entsprechende Gewichtsangpassung erfolgt. Im Rahmen der Simulationsstudie wird die Erweiterung von (4.13) auf den Delete-a-Group-Jackknife verwendet werden. Die Varianz wird hier geschätzt durch:

$$\hat{V}_{dgJK, strat}(\hat{\theta}) = \sum_{h=1}^H \frac{(1-f_h) \cdot (G_h-1)}{G_h} \sum_{g=1}^{G_h} \left(\hat{\theta}_{h,-g} - \overline{\hat{\theta}}_h \right)^2, \quad (4.17)$$

mit $\hat{\theta}_{h,-g} = T(y_{h,-g}, w_{h,-g})$. $y_{h,-g}$ ist der Vektor der Untersuchungsvariable in Schicht h ohne die Elemente der Gruppe g dieser Schicht. $w_{h,-g}$ ist ein Vektor, welcher die Gewichte $w_{h,j}$ enthält. Hinsichtlich der Bestimmung dieser Gewichte gilt für Element j bei Löschung der Gruppe g in Schicht h :

$$w_{h,j} = w_{hj} \cdot n_h / (n_h - n_{hg}). \quad (4.18)$$

n_{hg} ist die Anzahl der Elemente in Schicht h , welche sich in der gelöschten Gruppe g befinden. G_h ist die Anzahl der gebildeten Gruppen in Schicht h .

Wie beschrieben besteht der große Vorteil des Delete-a-Group-Jackknife in dem reduzierten Aufwand und den schnelleren Laufzeiten, insbesondere gegenüber dem Delete-1-Jackknife. Die Qualität des Verfahrens hängt entscheidend von der Anzahl der gebildeten Gruppen ab, welche die Anzahl der Replikationen bestimmen. Eine kleine Gruppenanzahl führt zu schnelleren Laufzeiten, aber auch zu einer größeren Variabilität der Schätzung. Bei vielen Anwendungen ist es häufig nicht so einfach, die optimale Anzahl an Gruppen zu bestimmen. Des Weiteren kann die Qualität der Methode leiden, wenn die Daten Klumpenstrukturen aufweisen. Werden ganze Cluster gelöscht und unterscheiden sich diese sehr stark, kann die Varianzschätzung mittels dieser Methode sehr variabel werden. Als Beispiel können Einheiten übergeordneter Stufen bei mehrstufigen Stichprobendesigns genannt werden.

4.3.2 Jackknife bei fehlenden Werten und Imputation

Ein bekannter Artikel, welcher sich mit der Anwendung des Delete-1-Jackknife im Falle fehlender Werte und Imputation beschäftigt, ist derjenige von RAO und SHAO (1992).

Im Rahmen von Nearest-Neighbour-Imputation sind für diese Resampling-Methode Besonderheiten gegeben. Diese werden in CHEN und SHAO (2001) näher erläutert und hier werden auch Modifikationen des Delete-1-Jackknife speziell für diese Imputationsmethode präsentiert. Im Rahmen der Simulationsstudie dieser Arbeit wird der Delete-1-Jackknife aufgrund seines hohen Aufwandes bei großen Stichprobenumfängen nicht berücksichtigt. Für detaillierte Erläuterungen wird daher auf diesen Artikel verwiesen.

Dieser Abschnitt stellt insbesondere auf den Delete-a-Group-Jackknife ab. Die Erkenntnisse aus RAO und SHAO (1992) und SHAO (2002) für den Delete-1-Jackknife können aber auch auf diese Methode übertragen werden. Auch für den Delete-a-Group-Jackknife gilt, dass durch die Löschung ganzer Gruppen, die Substichprobengröße kleiner als die ursprüngliche Stichprobengröße ist. Eine einfache Reimputation der imputierten Werte wie in Abschnitt 4.1.2 beschrieben, würde im Falle von zufälligen Imputationsverfahren zu einer Überschätzung der Varianz führen. In diesem Fall müssen die imputierten Werte in jeder Substichprobe, also die Stichprobenwerte ohne die gelöschten Elemente, wieder durch Formel (4.3) adjustiert werden. Grundlage dieser Adjustierungen ist der imputierte um die Gruppe g reduzierte Vektor y_{-g}^I bzw. $y_{h,-g}^I$, wobei bei letzterem beachtet werden muss, dass die Schichten nicht unbedingt die Imputationsklassen darstellen müssen. Aus Gründen der Übersichtlichkeit wird die Anwendungen nachfolgend am Vektor y_{-g} demonstriert. Nach der Anwendung von (4.3) ergibt sich die adjustierte imputierte Variable $y_{-g,adjust}$. Für den Punktschätzer in jeder Replikation für den Delete-a Group-Jackknife gilt:

$$\hat{\theta}_{-g} = T(y_{-g,adjust}, w_{-g}). \quad (4.19)$$

Die Varianzschätzung erfolgt anschließend wie im Falle vollständiger Beobachtungen nach (4.16). Im stratifizierten Fall würde entsprechend auf (4.17) zurückgegriffen werden.

Im Falle deterministischer Imputationsverfahren können sowohl die Adjustierung imputierter Werte wie auch die Reimputation angewendet werden. Beide führen dann zu gleichen Ergebnissen. Bei Anwendung der Reimputation würde die reimputierte Variable $y_{-g,reimp}$ resultieren und entsprechend verwendet werden:

$$\hat{\theta}_{-g} = T(y_{-g,reimp}, w_{-g}). \quad (4.20)$$

Auch hier erfolgt die Varianzschätzung nach (4.16).

4.4 Bootstrap

Die gängigste Variante eines Bootstrap ist der Monte-Carlo-Bootstrap. Im Falle einer einfachen Zufallsstichprobe werden aus der ursprünglichen Stichprobe n Einheiten $y_1^*, y_2^* \dots y_n^*$ in Höhe der Stichprobengröße nach dem Modell mit Zurücklegen gezogen. Die dazugehörigen Realisationen der entsprechenden Untersuchungsvariable y^* wird Bootstrap-Stichprobe genannt. Dieser Vorgang der Substichprobenziehung wird B mal wiederholt und für jede Substichprobe die interessierende Statistik $\hat{\theta}_b^*$ berechnet. Die Varianz wird berechnet durch

$$\hat{V}_{boot,MC}(\hat{\theta}) = \frac{1}{B-1} \cdot \sum_{b=1}^B \left(\hat{\theta}_b^* - \frac{1}{B} \cdot \sum_{v=1}^B \hat{\theta}_v^* \right)^2, \quad (4.21)$$

mit $\hat{\theta}_b^* = T(y^*, w_{boot}^*)$. w_{boot}^* stellt die Designgewichte der ursprünglichen Stichprobe der in die Bootstrap-Stichprobe gelangten Elemente dar (vgl. z.B. MÜNNICH, 2008, S. 326, vgl. SHAO und TU, 1995, S. 11 ff.).

Für geschichtete Stichprobendesigns erfolgt die Substichprobenziehung unabhängig in jeder Schicht (vgl. DAVISON und SARDY, 2004a, S. 21, DAVISON und SARDY, 2007, S. 374).

4.4.1 Monte-Carlo-Bootstrap bei fehlenden Werten und Imputation

Die wohl bekannteste Variante eines Monte-Carlo-Bootstrap bei fehlenden Werten bzw. Imputation ist der Bootstrap von SHAO und SITTER (1996). Wie beim Bootstrap üblich, werden hier n Einheiten mit Zurücklegen aus der imputierten Stichprobe mit den Werten der interessierenden Variable y^I gezogen. Es resultiert wieder die Bootstrap-Stichprobe mit y^{I*} mit den dazugehörigen Hilfsinformationen. Die Größe der Substichprobe entspricht der ursprünglichen Stichprobengröße. Zur Reproduktion des ursprünglichen Imputationsprozesses kann sowohl eine Reimputation, wie auch die Adjustierung der imputierten Werte in jeder Substichprobe erfolgen. Dies gilt auch für zufällige Imputationsverfahren. SHAO und SITTER (1996) benutzen in ihrem Artikel die Reimputation innerhalb jeder Substichprobe, gemäß dem folgenden Vorgehen:

Es wird davon ausgegangen dass z_i^* den entsprechen Response-Indikator für die Bootstrap-Stichprobe beschreibt, mit $z_i^* = 0$, falls die Einheit i der Bootstrap-Stichprobe einen fehlenden Wert besitzt und zuvor imputiert wurde ($y_M^* = y_i^* \in M^*$) und 1, wenn dieser beobachtet worden ist ($y_R^* = y_{hi}^* \in R^*$). Um die Imputation bei der Varianzschätzung zu berücksichtigen, werden die fehlenden Werte y_M^* der Bootstrap-Stichprobe in jeder Substichprobe auf Basis der beobachtbaren Werte y_R^* reimputiert. Um y_{reimp}^* zu generieren, wird das gleiche Imputationsverfahren benutzt, welches für die ursprüngliche Stichprobe verwendet wurde. Die interessierende Statistik wird anschließend durch $\hat{\theta}_b^* = T(y_{reimp}^*, w_{boot}^*)$ berechnet und die Varianzschätzung erfolgt nach (4.21). Nach SHAO (2002, S. 309) ist eine Anpassung der imputierten Werte nach (4.3) aus den genannten Gründen nicht notwendig.

Anders verhält es sich, wenn geschichtete Designs vorliegen. Nach RAO und WU (1988) müssen bei kleinen Stichprobenumfängen statt $n_h n_h - 1$ Einheiten mit Zurücklegen unabhängig in jeder Schicht aus der ursprünglichen Stichprobe gezogen werden. Damit entspricht die Substichprobengröße bei geschichteten Designs nicht mehr derjenigen der ursprünglichen Stichprobe. Im Falle zufälliger Imputationsverfahren können daher bei einer einfachen Reimputation in jeder Substichprobe deutliche Überschätzungen der Varianz auftreten. Die Adjustierung der imputierten Werte nach (4.3) ist hier sinnvoll (vgl. SHAO, 2002, S. 310). Im Rahmen des Shao und Sitter-Bootstrap ist damit das Stichprobendesign von entscheidender Bedeutung, ob eine Reimputation oder Adjustierung der imputierten Werte angewendet werden kann. Wie zuvor beschrieben, spielt es bei einer einfachen Zufallsstichprobe keine Rolle, welche der beiden Methoden verwendet wird. Liegen hingegen stratifizierte Designs oder Strukturen vor, wie z.B. auf der zweiten Stufe oder darauf folgenden Stufen eines mehrstufigen Stichprobendesigns, muss die Adjustierung der imputierten Werte innerhalb jeder Substichprobe des Shao und Sitter-Bootstrap erfolgen. Der Shao und Sitter-Bootstrap weist einige Nachteile auf. Zum einen kann er auch mit deutlichen Überschätzungen der Varianz einhergehen. Der Monte-Carlo-Bootstrap im Allgemeinen berücksichtigt keine Endlichkeitskorrektur für den Fall des Ziehens ohne Zurücklegen. Diese ist bei großen Auswahlsätzen relevant. Daher kann auch der Shao und Sitter-Bootstrap in solchen Fällen zu deutlichen Verzerrungen führen. Hier müssen alter-

native Bootstrap-Varianten gefunden werden. Zum anderen kann er nach SAIGO et al. (2001) zu deutlichen Unterschätzungen der Varianz führen, im Falle kleiner Stichprobenumfänge innerhalb der Schichten. Der Grund liegt auch hier darin, dass die Größe der Substichprobe in den verschiedenen Schichten in Höhe von $n_h - 1$ nicht mit der Größe der ursprünglichen Stichprobe übereinstimmt. Aus diesem Grund schlagen sie eine weitere Modifikation des Bootstrap vor, welche sie Repeated-Half-Sample-Bootstrap nennen.

4.4.2 Repeated-Half-Sample-Bootstrap

Hier werden bei Annahme einer gerade Anzahl von Einheiten in den Schichten anstelle von $n_h - 1$ Einheiten $n_h^* = n_h/2$ für die Substichprobe gezogen, allerdings im Modell ohne Zurücklegen. Anschließend werden diese Einheiten verdoppelt, wodurch eine Substichprobengröße von n_h erreicht wird (vgl. SAIGO et al., 2001, S. 191 f.). Im Falle einer ungeraden Stichprobengröße in den Schichten schlagen SAIGO et al. (2001) eine modifizierte Version dieses Bootstrap vor.

In diesem Zusammenhang gehen sie von zwei Überlegungen aus:

1. Werden $n_h^* = (n_h - 1)/2$ ohne Zurücklegen gezogen und alle Einheiten verdoppelt, resultieren nur $n_h - 1$ Einheiten. Wird nun zufällig von diesen Einheiten eine weitere gezogen, um die gewünschte Stichprobengröße zu erhalten, wird mit dem Repeated-Half-Sample-Bootstrap eine Varianz bei Vorliegen eines Totalwertschätzers geschätzt, welche $V^*(\tau) = \sum_{h=1}^H (n_h + 3) \cdot s_h^2/n_h^2$ beträgt.
2. Werden hingegen $n_h^* + 1$ Einheiten gezogen und repliziert, resultieren $n_h + 1$ Einheiten. Wird davon eine zufällig gelöscht, liegt die Varianz bei $V^*(\tau) = \sum_{h=1}^H (n_h + 1) \cdot s_h^2/n_h^2$.

Nach SAIGO et al. (2001) wird ein unverzerrter Varianzschätzer dann erreicht, wenn die erste Methode mit einer Wahrscheinlichkeit von $1/4$ und die zweite Methode mit einer Wahrscheinlichkeit von $3/4$ benutzt wird. Das weitere Vorgehen entspricht dem Shao und Sitter-Bootstrap (vgl. SAIGO et al., 2001, S. 191 f.). Wie in SAIGO et al. (2001) beschrieben, wird dadurch ein approximativ unverzerrter Bootstrap mit Substichprobengröße in Höhe der ursprünglichen Stichprobengröße erreicht. Dadurch ist es bei dieser Version des Bootstrap auch im Falle eines zufälligen Imputationsverfahrens möglich, eine Reimputation in jeder Substichprobe durchzuführen, ohne dass Überschätzungen der Varianz auftreten. Hier spielt es keine Rolle, ob die Reimputation oder die Adjustierung der imputierten Werte in jeder Substichprobe zur Berücksichtigung der durch die Imputation entstehende Varianz gewählt wird. Es können damit die Vorteile der Reimputation wie z.B. die einfachere Umsetzbarkeit ausgenutzt werden. Wird diese gewählt, resultiert die reimputierte nach dem Vorgehen des Repeated-Half-Sample-Bootstrap erhaltene interessierende Variable $y_{repeat, reimp}^*$ mit dem dazu gehörenden Gewichtevektor w_{repeat}^* . Dieser beinhaltet die Designgewichte der ursprünglichen Stichprobe dieser Einheiten. Hier muss keine Anpassung der Gewichte erfolgen, da die Substichprobengröße mit der ursprünglichen Stichprobengröße übereinstimmt. Der Punktschätzer wird berechnet durch $\hat{\theta}_b^* = T(y_{repeat, reimp}^*, w_{repeat}^*)$. Das weitere Vorgehen ist mit dem des Shao und Sitter-Bootstrap identisch und die Varianzschätzung erfolgt nach (4.21).

Aber auch der Repeated-Half-Sample-Bootstrap besitzt einen entscheidenden Nachteil. Wie der Shao und Sitter-Bootstrap berücksichtigt diese Version des Bootstrap keine Endlichkeitskorrektur. Dies kann im Falle großer Ausfallsätze innerhalb der Schichten $f_h = n_h/N_h$ zu deutlichen Überschätzungen der Varianz führen. Diese Problematik trat

auch schon für die Random-Group-Methode sowie den Delete-a-Group-Jackknife auf. Wie beschrieben wird diese Integration komplexer, wenn z.B. bei stratifizierten Designs unterschiedliche Auswahlsätze in den Schichten vorliegen. Zwei solche Methoden, welche in der Lage sind, Endlichkeitskorrekturen auch bei variierenden Auswahlätzen zu berücksichtigen, sind der Rescaling-Bootstrap ohne Zurücklegen und die Methode des Balanced-Repeated-Replication. Diese werden in den nachfolgenden Abschnitten näher erläutert.

4.5 Rescaling-Bootstrap ohne Zurücklegen

Eine Möglichkeit zur Behebung des beschriebenen Problems der Berücksichtigung der Endlichkeitskorrektur ist eine weitere Version des Bootstrap, der Rescaling-Bootstrap ohne Zurücklegen, welcher von CHIPPERFIELD und PRESTON (2007, S. 169f.) entwickelt wurde. Dieses Verfahren stellt eine Weiterentwicklung des von RAO und WU (1988) vorgeschlagenen Rescaling-Bootstrap dar, welcher zunächst in Abschnitt 4.5.1 kurz beschrieben wird. Im Rahmen dieser Arbeit wird jedoch größten Teils auf die Weiterentwicklung von CHIPPERFIELD und PRESTON (2007) abgestellt. Seine Übertragung auf mehrstufige Stichprobendesigns wird in PRESTON (2009, p. 228 ff.) gegeben. Sowohl die ein- wie auch die mehrstufige Variante des Rescaling-Bootstrap ohne Zurücklegen werden in den nächsten Abschnitten vorgestellt. Beide Varianten gehen von vollständigen Beobachtungen aus. Anschließend wird in dieser Arbeit ein Ansatz vorgeschlagen, wie diese beiden Varianten auf Fälle des Vorliegens fehlender Werte und ihrer Kompensation über Imputation übertragen werden können.

4.5.1 Ursprüngliche Version nach Rao und Wu

In seiner ursprünglichsten Version stellt der Rescaling-Bootstrap eine Resampling-Methode dar, welche noch auf dem Modell mit Zurücklegen basiert. Genauer gesagt, werden im Falle einer geschichteten Zufallsstichprobe eine bestimmte Anzahl von n_h^* Einheiten aus der ursprünglichen Stichprobe mit Zurücklegen gezogen. RAO und WU (1988) schlagen hier die Verwendung von

$$n_h^* = \left\lfloor \frac{(1 - f_h) \cdot (n_h - 2)^2}{1 - 2 \cdot f_h \cdot (n_h - 1)} \right\rfloor \quad (4.22)$$

vor. Anschließend erfolgt eine Anpassung der für die Substichprobe gezogenen Werte y_{hi}^* durch:

$$\bar{y}_{hi} = \bar{y}_h + (y_{hi}^* - \bar{y}_h) \cdot \sqrt{C_h}, \quad (4.23)$$

mit $C_h = \frac{n_h^* \cdot (1 - f_h)}{(n_h - 1)}$,

wobei \bar{y}_h den Schichtmittelwert darstellt. Die Designgewichte werden berechnet durch $w_{hi}^* = N_h/n_h^*$. Für die interessierende Statistik ergibt sich $\hat{\theta}_b^* = T(y_h^*, w_h^*)$, wobei y_h^* und w_h^* die entsprechenden Vektoren von y_{hi}^* und w_{hi}^* sind (vgl. RAO und WU, 1988, S. 232, DAVISON und SARDY, 2004a, S. 22, MASHREGHI et al., 2014, S. 147).

Dieser ursprüngliche Rescaling-Bootstrap besitzt eine Reihe von Nachteilen. Das größte Problem liegt in der Bestimmung der Substichprobengröße. So kritisieren DAVISON und

SARDY (2004a), dass deren Bestimmung nach (4.22) unter bestimmten Parameterkonstellationen zu $n_h^* > n_h$ führen kann. Dadurch kann die Berechnung des Rescaling-Punktschätzers nicht mehr möglich sein. MASHREGHI et al. (2014) schlagen als Alternative die Verwendung von $n_h - 3$ vor. Aber auch diese Wahl kann Probleme mit sich bringen, wie die Simulationsstudie in Abschnitt 5.2 zeigen wird. PRESTON (2009) kritisiert zudem, dass die Reskalierung auf die Beobachtungswerte durchgeführt wird. Dadurch ist seine Anwendung nur bei glatten Statistiken möglich. Ein letzter Kritikpunkt wird in SAIGO et al. (2001) gegeben. Nach ihrer Auffassung liefert die ursprüngliche Version des Rescaling-Bootstrap keine korrekten Schätzer im Falle von imputierten Daten.

RAO et al. (1992) modifizierten diese Methode und führten die Reskalierungen auf die Designgewichte und nicht mehr auf die Beobachtungswerte durch. Im Rahmen dieser Arbeit wird wie bereits erläutert zu einem überwiegenden Teil auf eine weitere Modifikation des Rescaling-Bootstrap abgestellt, welcher in CHIPPERFIELD und PRESTON (2007) vorgeschlagen wird und den beschriebenen Nachteilen begegnen soll. Zusätzlich besitzt diese Methode große Vorteile bei der Anwendung mehrstufiger Stichprobendesigns. Der große Unterschied im Vergleich zu der beschriebenen originären Version des Rescaling-Bootstrap ist jedoch die Verwendung des Modells ohne Zurücklegen für die Substichprobenziehung. Daher wird diese Version im Folgenden auch oft als Rescaling-Bootstrap ohne Zurücklegen bezeichnet.

4.5.2 Allgemeines Vorgehen

Auch beim Rescaling-Bootstrap ohne Zurücklegen wird eine Substichprobe nicht in Höhe der ursprünglichen Stichprobengröße gezogen. Im Falle der einfachen Zufallsstichprobe werden von den n Elementen der betrachteten Stichprobe $n^* = \lfloor n/2 \rfloor$ Elemente gezogen, aber nicht im Modell mit Zurücklegen sondern im Modell ohne Zurücklegen (vgl. hierzu und im Folgenden CHIPPERFIELD und PRESTON, 2007, S. 169f.). Bei vorausgehender Schichtung erfolgt dieser Vorgang unabhängig in jeder Schicht. Im Gegensatz z.B. zum Repeated-Half-Sample-Bootstrap werden diese gezogenen Einheiten aber nicht verdoppelt. Beim Rescaling-Bootstrap erfolgt im Falle der einfachen Zufallsstichprobe eine Adjustierung der Gewichte aller Einheiten der ursprünglichen Stichprobe durch:

$$w_i^* = \left(1 - \lambda + \lambda \cdot \frac{n}{n^*} \cdot \delta_i\right) \cdot w_i, \quad (4.24)$$

mit dem Gewichtevektor w_{Resc}^* aller Beobachtungen, wobei δ_i den Wert 1 annimmt, wenn das betrachtete Element für die Substichprobe gezogen wurde und ansonsten den Wert 0.

λ ist definiert durch $\lambda = \sqrt{n^* \cdot \frac{(1-f)}{(n-n^*)}}$. Wie zu sehen ist, wird bei der Berechnung des

Rescaling-Gewichtes die Endlichkeitskorrektur $(1-f)$ berücksichtigt. Dadurch kann eine Überschätzung der Varianz bei einem großen Auswahlatz vermieden werden. Und dies ist der große Vorteil gegenüber den zuvor vorgestellten Bootstrap-Verfahren wie des Shao und Sitter-Bootstrap und des Repeated-Half-Sample-Bootstrap, wie sich später noch bei der Simulationsstudie zeigen wird.

Zur Berechnung des Bootstrap-Punktschätzers werden anschließend die Gewichte w_i^* verwendet. Hier gilt $\hat{\theta}_b^* = T(y_S, w_{Resc}^*)$. Wichtig ist an dieser Stelle noch einmal hervorzuheben, dass bei dieser Variante des Bootstrap alle Elemente der ursprünglichen Stichprobe y_S bei Berechnung der einzelnen für die Varianzschätzung relevanten Bootstrap-Punktschätzer berücksichtigt werden. Die in der Substichprobe enthaltenen Elemente, für welche $\delta_i = 1$ gilt, bekommen jedoch ein höheres Gewicht. Dies wird deutlich, wenn in

Formel (4.24) δ_i auf 0 bzw. 1 gesetzt wird. Dabei erhöht sich das Gewicht der gezogenen Elemente gegenüber den nicht gezogenen Elementen um den Term $\lambda \cdot \frac{n}{n^*}$. Dies resultiert aus der Tatsache, dass n, n^* und λ nur positive Werte annehmen können. Die für die Substichprobe gezogenen Elemente ($\delta_i = 1$) besitzen in diesem Zusammenhang im Vergleich zu den nicht gezogenen ein umso größeres Gewicht, je kleiner der Auswahlatz f ist. Jedoch gewinnen die nicht gezogenen Elementen mit größer werdendem Auswahlatz immer mehr an Bedeutung. Wird angenommen, dass der Schichtumfang n gerade ist und damit $n^* = n/2$, gilt für den zusätzlichen Term $\lambda \cdot \frac{n}{n^*} = 2 \cdot \sqrt{(1-f)}$. Mit großem f nähert sich dieser Term Null und die Gewichte w_i^* nach (4.24) unterscheiden sich kaum noch. Wie zuvor erfolgt die Varianzschätzung auf Basis der einzelnen aus den B Wiederholungen resultierenden Bootstrap-Punktschätzer nach Formel (4.21).

Im Falle der geschichteten Zufallsstichprobe ändert sich (4.24) zu:

$$w_{hi}^* = \left(1 - \lambda_h + \lambda_h \cdot \frac{n_h}{n_h^*} \cdot \delta_{hi} \right) \cdot w_{hi}, \quad (4.25)$$

mit dem Gewichtevektor $\lambda_h = \sqrt{n_h^* \cdot \frac{(1-f_h)}{(n_h - n_h^*)}}$ aller Beobachtungen. δ_{hi} nimmt auch hier den Wert 1 an wenn eine Einheit i der Schicht h für die Substichprobe gezogen wurde und ansonsten 0 (vgl. CHIPPERFIELD und PRESTON, 2007, S. 169 f.). Bezüglich der Einflussnahme für die Substichprobe gezogener und nicht gezogener Einheiten gelten die Erläuterungen aus dem Fall der einfachen Zufallsstichprobe. Die Einflussnahme der nicht für die Substichprobe gezogenen Einheiten wächst mit steigendem f_h .

4.5.3 Rescaling-Bootstrap ohne Zurücklegen bei mehrstufigen Stichprobendesigns

Neben seiner Vorteile bei höheren Auswahlätzen durch die Berücksichtigung der Endlichkeitskorrektur bei der Berechnung der Rescaling-Gewichte besitzt der Rescaling-Bootstrap einen weiteren Vorteil. Im Falle des Vorliegens mehrstufiger Stichprobendesigns kann die Methode unkompliziert auf mehrere Stufen erweitert werden. Dies geschieht über die Adjustierung der Designgewichte auf der entsprechenden Stufe.

Diese Erweiterung des Rescaling-Bootstrap auf mehrstufige Stichprobendesigns wird in PRESTON (2009) beschrieben, wobei dieser ein dreistufiges Design mit einer stratifizierten ersten Stufe verwendet. Im Rahmen dieses Abschnittes wird auf ein zweistufiges Stichprobendesign abgestellt werden, ebenfalls mit einer Schichtung auf der ersten Stufe. Das grundsätzliche Vorgehen des Rescaling-Bootstrap besteht im mehrstufigen Fall darin, auf jeder Stufe für die Substichprobe die Hälfte der Einheiten ohne Zurücklegen zu ziehen und die Designgewichte dementsprechend anzupassen. Im Detail wird der Rescaling-Bootstrap ohne Zurücklegen wie folgt umgesetzt. Zunächst werden im Rahmen der Substichprobenbildung aus den l_h PSU der betrachteten Stichprobe $l_h^* = \lfloor l_h/2 \rfloor$ PSU zufällig ohne Zurücklegen ausgewählt. Es erfolgt eine Berechnung der Rescaling-Gewichte w_{hd}^* auf dieser Stufe gemäß:

$$w_{hd}^* = \left(1 - \lambda_h + \lambda_h \cdot \frac{l_h}{l_h^*} \cdot \delta_{hd} \right) \cdot w_{hd}. \quad (4.26)$$

Für λ_h gilt $\lambda_h = \sqrt{l_h^* \cdot \frac{(1-f_h)}{(l_h - l_h^*)}}$ und δ_{hd} ist eine Indikatorvariable, welche den Wert 1 annimmt, wenn PSU d in Schicht h gezogen wird und ansonsten gilt $\delta_{hd} = 0$.

Auf der zweiten Stufe wird dieser Vorgang entsprechend durchgeführt. Demnach werden von den n_{hd} USU in der d -ten PSU der ursprünglichen Stichprobe $n_{hd}^* = \lfloor n_{hd}/2 \rfloor$ USU zufällig ohne Zurücklegen ausgewählt. Dieser Vorgang wird unabhängig von der Substichprobenziehung der ersten Stufe gemacht, d.h. die Substichprobenziehung der zweiten Stufe findet aus allen l_h PSU der ursprünglichen Stichprobe statt und nicht aus den l_h^* Einheiten der Substichprobe.

Die Rescaling-Gewichte auf dieser Stufe sind definiert durch:

$$w_{hdi}^* = w_{hdi} \cdot \frac{w_{hd}}{w_{hd}^*} \cdot \left(1 - \lambda_h + \lambda_h \cdot \frac{l_h}{l_h^*} \cdot \delta_{hd} - \lambda_{hd} \cdot \sqrt{\frac{l_h}{l_h^*}} \cdot \delta_{hd} + \lambda_{hd} \cdot \sqrt{\frac{l_h}{l_h^*}} \cdot \delta_{hd} \cdot \frac{n_{hd}}{n_{hd}^*} \cdot \delta_{hdi} \right), \quad (4.27)$$

wobei λ_{hd} definiert ist durch $\lambda_{hd} = \sqrt{n_{hd}^* \cdot f_h \cdot \frac{(1-f_{hd})}{(n_{hd} - n_{hd}^*)}}$ und δ_{hdi} besitzt den Wert 1 wenn die USU i in PSU d in Schicht h gezogen wird und ansonsten 0.

Die neuen Gewichte errechnen sich aus:

$$w_i^* = w_{hd}^* \cdot w_{hdi}^*, \quad (4.28)$$

mit dem Gewichtevektor w_{Resc}^* (vgl. PRESTON, 2009, S. 228 f.).

Auch hier gilt tendenziell die Aussage, dass bei geringen Auswahlätzen auf den jeweiligen Stufen die Gewichte der für die Substichprobe gezogenen Einheiten im Vergleich zu den nicht gezogenen Einheiten deutlich größer sind als im Fall von großen Auswahlätzen. Hier nimmt der Einfluss der nicht gezogenen Einheiten deutlich zu. Im Vergleich zum einstufigen Design können hier jedoch weniger eindeutige Aussagen getroffen werden, da mehr Parameter eine Einflussnahme auf die Gewichtsangpassung besitzen. So beeinflussen zum Beispiel die Auswahlätze der ersten Stufe auch die Gewichtsangpassung auf der zweiten Stufe. Dies muss bei der folgenden Interpretation beachtet werden.

Für die Gewichte aus (4.26) gelten die getätigten Erläuterungen aus dem einstufigen Design, aber bezüglich der PSU als relevante Einheit. Bei kleinen Auswahlätzen von PSU innerhalb der Schichten f_h sind die angepassten Gewichte der für die Substichprobe gezogenen Einheiten ($\delta_{hd} = 1$) deutlich größer. Es kommt jedoch zu einer Annäherung der Gewichte mit steigendem f_h . Auf der zweiten Stufe wird die Analyse der Gewichte komplexer, da hier wie beschrieben auch die erste Stufe Einfluss besitzt. Bei einer geraden Anzahl an Stichprobenumfängen auf den Stufen kann (4.27) umgeformt werden zu:

$$w_{hdi}^* = w_{hdi} \cdot \frac{w_{hd}}{w_{hd}^*} \cdot \left(1 - \sqrt{(1-f_h)} + 2 \cdot \sqrt{(1-f_h)} \cdot \delta_{hd} - \sqrt{2} \cdot \sqrt{f_h \cdot (1-f_{hd})} \cdot \delta_{hd} + 2 \cdot \sqrt{2} \cdot \sqrt{f_h \cdot (1-f_{hd})} \cdot \delta_{hd} \cdot \delta_{hdi} \right). \quad (4.29)$$

Hierbei wird berücksichtigt, ob eine Einheit über ihre PSU auf der ersten Stufe für die Substichprobe gezogen worden ist ($\delta_{hd} = 1$). Wird eine Einheit auch auf der zweiten Stufe gezogen ($\delta_{hdi} = 1$), so kommt bei der Gewichtsadjustierung zusätzlich der Term

$2 \cdot \sqrt{2} \cdot \sqrt{f_h \cdot (1 - f_{hd})} \cdot \delta_{hd} \cdot \delta_{hdi}$ hinzu. Bei gegebenen f_h ist auch hier das Gewicht der für die Substichprobe gezogenen Einheiten ($\delta_{hd} = 1$ und $\delta_{hdi} = 1$) bei kleinen Auswahlsätzen f_{hd} deutlich größer. Bei großen Auswahlsätzen nimmt der Einfluss der auf der ersten aber nicht auf der zweiten Stufe gezogenen Einheiten ($\delta_{hd} = 1$ und $\delta_{hdi} = 0$) zu. An dem Term $2 \cdot \sqrt{2} \cdot \sqrt{f_h \cdot (1 - f_{hd})} \cdot \delta_{hd} \cdot \delta_{hdi}$ lässt sich des Weiteren auch der folgende Sachverhalt deuten. Werden Einheiten auf der ersten Stufe nicht für die Substichprobe gezogen, jedoch auf der zweiten Stufe ($\delta_{hd} = 0$ und $\delta_{hdi} = 1$), so erfolgt eine Anpassung der Designgewichte dieser Einheiten um den gleichen Term, wie für Einheiten, welche auf keiner der beiden Stufen gezogen werden ($\delta_{hd} = 0$ und $\delta_{hdi} = 0$). Der Indikator δ_{hdi} taucht in (4.27) bzw. (4.29) nur in diesem Term auf, welcher verschwindet im Fall von $\delta_{hd} = 0$. Damit kann der Fall $\delta_{hd} = 0$ und $\delta_{hdi} = 1$ auch als nicht für die Substichprobe gezogen betrachtet werden. Sowohl der Fall $\delta_{hd} = 0$ und $\delta_{hdi} = 1$ wie auch der Fall $\delta_{hd} = 0$ und $\delta_{hdi} = 0$ können mit großen f_{hd} an Bedeutung gewinnen. Bei gegebenen f_h und immer größer werdender f_{hd} konvergieren die beiden letzten Terme aus (4.29) gegen Null. Liegen zusätzlich große Auswahlsätze in den Schichten f_h vor, wird auch der Term $2 \cdot \sqrt{(1 - f_h)} \cdot \delta_{hd}$ kleiner. Die Gewichtsadjustierungen der für die Substichprobe gezogenen Einheiten auf der zweiten Stufe können sich unter diesen Umständen denjenigen der nicht gezogenen Einheiten annähern. Diese beträgt $\frac{w_{hd}}{w_{hd}^*} \cdot (1 - \sqrt{(1 - f_h)})$. Auf der ersten Stufe wurde schon erläutert, dass die Unterschiede zwischen den Gewichten bei großen f_h geringer sind.

Die auf diesem Wege errechneten Gewichte werden beim Bootstrap-Punktschätzer in (4.21) verwendet. Es gilt: $\hat{\theta}_b^* = T(y_s, w_{Resc}^*)$. Wie im einstufigen Fall werden für die Berechnung des Punktschätzers in der Substichprobe alle Einheiten der ursprünglichen Stichprobe y_s benutzt.

Die Wahl von $l_h^* = \lfloor l_h/2 \rfloor$, und $n_{hd}^* = \lfloor n_{hd}/2 \rfloor$ ist nach PRESTON (2009, S. 229) die optimale Substichprobengröße. Dies hat zudem den Vorteil, dass negative Gewichte vermieden werden (vgl. PRESTON, 2009, S. 229).

Das beschriebene Vorgehen des Rescaling Bootstrap wird auch bei mehr als zwei Stufen beibehalten. Die Gewichtsadjustierung auf folgenden Stufen wird rekursiv fortgesetzt (vgl. hierzu PRESTON, 2009, S. 229, insbesondere die auf dieser Seite angegebene Formel zur Erweiterung des Rescaling-Bootstrap auf zusätzliche Stufen).

Und dies ist ein weiterer großer Vorteil dieses Verfahrens. Es muss nicht auf jeder Stufe ein neuer Varianzschätzer, wie in Abschnitt 4.1.3 beschrieben, konzipiert werden. Auf jeder Stufe müssen nur die Gewichte angepasst werden, um die Varianz auf den verschiedenen Stufen zu berücksichtigen. Damit erfolgt die Berücksichtigung der verschiedenen Stufen schon bei der Berechnung des Rescaling-Punktschätzers. Dadurch ist der Rescaling-Bootstrap auch bei Stichprobendesigns mit deutlich mehr als drei Stufen unkompliziert und programmieretechnisch einfach anwendbar. Zudem können die Laufzeiten für solche Stichprobendesigns im Vergleich zu anderen Resampling-Methoden deutlich geringer sein. Die Gewichtsadjustierung kann zwar sehr aufwendig werden. Im Falle von zwei Stufen ist sie noch übersichtlich, aber bei deutlich mehr als zwei Stufen kann sich dies schnell ändern. Aber dadurch, dass der Rescaling-Bootstrap nicht auf jeder Stufe vollständig und separat angewendet wird, sondern die einzelnen Stufen in der Berechnung des Punktschätzers berücksichtigt werden, können Laufzeiten reduziert werden.

4.5.4 Rescaling-Bootstrap ohne Zurücklegen bei fehlenden Werten und Imputation

Der zuvor beschriebene Rescaling-Bootstrap basiert sowohl im einstufigen wie auch im mehrstufigen Design auf vollständigen Daten. Ein Ziel dieser Arbeit und insbesondere dieses und des nächsten Abschnittes ist es, diese Methode auf den Fall fehlender Werte und Imputation hin zu modifizieren. Die Ansätze zur Übertragung von klassischen Resampling-Methoden auf den Fall von Imputation, wie sie in der Literatur angewandt werden, sind hier, wie noch zu sehen sein wird, nur begrenzt übertragbar, da sich der Rescaling-Bootstrap ohne Zurücklegen von den klassischen Methoden unterscheidet. Hier muss ein neuer Ansatz entwickelt werden, welcher in den folgenden Ausführungen sowohl für das ein- wie auch mehrstufige Design theoretisch hergeleitet wird. Zunächst wird das einstufige Design betrachtet. In diesem Abschnitt wird aus Gründen der Übersichtlichkeit von der einfachen Zufallsstichprobe ausgegangen. Das Verfahren kann jedoch auch für die geschichtete Zufallsstichprobe angewendet werden, wie die Simulationsstudie zeigen wird. Um die Imputation zu berücksichtigen, muss zunächst gemäß den Ausführungen in Abschnitt 4.1.2 überlegt werden, ob in der Substichprobe eine Reimputation fehlender Werte möglich ist oder eine Adjustierung der imputierten Werte gemäß (4.3) stattfinden muss. Und hier ist entscheidend, dass für jede Replikation $\lfloor n/2 \rfloor$ Einheiten und nicht n Einheiten gezogen werden. Dadurch würde eine einfache Reimputationen bei zufälligen Imputationsverfahren zu Überschätzungen der Varianz führen. Daher wäre es hier sinnvoll, die Adjustierung der imputierten Werte in jeder Replikation gemäß Formel (4.3) zu verwenden. Die Simulationsstudie wird später noch zeigen, dass dies bei zufälligen Imputationsverfahren das richtige Vorgehen ist. Bei deterministischen Imputationsverfahren führen beide Methoden zu gleichen Ergebnissen und auch eine Reimputation kann hier angewendet werden.

Problematisch stellt sich für den Rescaling-Bootstrap ohne Zurücklegen dar, dass im Vergleich zu anderen klassischen Resampling-Verfahren zwar nur die Hälfte der Einheiten gezogen werden, jedoch zur Berechnung des Punktschätzers alle Einheiten eingehen. Über die Gewichtung in (4.24) wird jedoch ein größerer Einfluss der Einheiten, welche für die Substichprobe ausgewählt wurden ($\delta_i = 1$), zumindest bei einem kleinen Auswahlsatz, gewährleistet. Hier bekommen Einheiten, welche nicht gezogen wurden ($\delta_i = 0$) ein deutlich kleineres Gewicht. Mit zunehmendem Auswahlsatz wird das Gewicht dieser nicht gezogenen Einheiten im Vergleich zu den gezogenen Einheiten immer größer und sie besitzen einen größeren Einfluss bei der Berechnung des Punktschätzers. Und für solche Fälle ist es wichtig zu wissen, wie eine Reimputation bzw. eine Adjustierung der imputierten Werte durchgeführt werden kann.

Hier wäre eine erste Überlegung, die Reimputation bzw. Adjustierung in jeder Substichprobe auf alle imputierten Werte, unabhängig ihrer Ziehung für die Substichprobe, durchzuführen. Der Hintergrund für dieses Vorgehen besteht darin, dass für die Berechnung des Rescaling-Punktschätzers alle Einheiten der ursprünglichen Stichprobe berücksichtigt werden. Bei deterministischen Imputationsverfahren würden in jeder Substichprobe exakt die gleichen Imputationswerte wie die der ursprünglichen Stichprobe resultieren und über die Substichproben hinweg würde keine Variation entstehen. Dieses Vorgehen wäre nicht geeignet.

Im Rahmen dieser Arbeit werden daher zwei Modifikationen des Rescaling-Bootstrap ohne Zurücklegen vorgeschlagen.

- *Modifikation 1:* Bei der ersten Modifikation werden nur die imputierten Werte der für die Substichprobe gezogenen Einheiten, also die Einheiten bei denen δ_i den Wert

1 annimmt und $i \in M$, reimputiert bzw. adjustiert. Dieser Vorgang geschieht auf Basis der für die Substichprobe gezogenen beobachteten Einheiten ($\delta_i = 1$ und $i \in R$). Dies würde den Gedanken eines einfachen Shao und Sitter-Bootstrap widerspiegeln. Durch diese Vorgehensweise wird möglicherweise die Varianz nicht richtig erfasst. Der Grund liegt darin, dass zur Berechnung des Punktschätzers des Rescaling-Bootstrap zum einen imputierte Werte eingehen, welche auf Basis der für die Substichprobe gezogenen beobachteten Einheiten reimputiert oder adjustiert worden sind. Hierbei handelt es sich um die Einheiten mit fehlendem Wert und $\delta_i = 1$. Wurden die Einheiten für die Substichprobe nicht gezogen und $\delta_i = 0$, gehen in dieser Modifikation für fehlende Werte jene Imputationswerte ein, welche auf Basis der ursprünglichen Stichprobe imputiert worden sind. Bei Betrachtung aller Substichproben wird eine fehlende Einheit in manchen Substichproben mit dem Imputationswert der ursprünglichen Stichprobe, in anderen Substichproben mit dem reimputierten bzw. adjustierten Wert eingehen, also Werte deren Berechnungen auf unterschiedlichen Datengrundlagen basieren. Über die Substichproben hinweg entsteht dadurch zu viel Variation, welche in dieser Form im ursprünglichen Imputationsprozess nicht entstanden ist. Dadurch können sich Überschätzungen der Gesamtvarianz ergeben und damit Verzerrungen auftreten. Resampling-Verfahren sollen das ursprüngliche Stichproben- und Imputationsschema möglichst gut widerspiegeln, was hier dann nicht der Fall wäre. Im originären Imputationsprozess gab es die beiden verschiedenen Datengrundlagen zur Imputation in dieser Form nicht.

- *Modifikation 2:* Daher wird noch eine weitere Modifikation vorgeschlagen, bei der zusätzlich auch die Imputationswerte der nicht für die Substichprobe gezogenen Einheiten, also mit $\delta_i = 0$ reimputiert bzw. nach (4.3) adjustiert werden. Damit gehen Einheiten mit fehlendem Wert über die Substichproben hinweg immer mit dem reimputierten bzw. adjustierten Wert ein. Imputationswerte der ursprünglichen Stichprobe kommen in den Substichproben nicht mehr vor. Dadurch wird die beschriebene ungewünschte Variation vermieden, welche im ursprünglichen Imputationsprozess nicht aufgetreten ist.

Die getätigten Ausführungen spielen jedoch, wie erläutert, nur bei einem großen Auswahlatz ein Rolle. Bei einem kleinen Auswahlatz besitzen die Elemente mit $\delta_i = 0$ kaum Einfluss bei der Berechnung des Punktschätzers des Rescaling-Bootstrap und zwischen den Modifikationen bestehen kaum Unterschiede. In diesem Fall fallen bei der Berechnung der Punktschätzer der Substichproben hauptsächlich die Elemente mit $\delta_i = 1$ ins Gewicht. Die Reimputationen bzw. Adjustierungen mit großem Einfluss auf die Varianzschätzung sind bei beiden Modifikationen damit hier bei den fehlenden Werten mit $\delta_i = 1$ geben. Zur Umsetzung der beiden Modifikationen des Rescaling-Bootstrap ergibt sich zusammenfassend das folgende Vorgehen:

Bei beiden Modifikation werden in jeder Substichprobe die für die Substichprobe gezogenen fehlenden Werte betrachtet, also die Einheiten $i \in M$ bei denen δ_i den Wert 1 annimmt. Bei der Adjustierung der imputierten Werte werden die beiden Erwartungswerte aus (4.3) auf Basis der entsprechenden beobachteten Werte $i \in R$ mit $\delta_i = 1$ berechnet. Bei der Reimputation sind diese Einheiten die Grundlage zur Wiederholung des Imputationsschemas der ursprünglichen Stichprobe. Im Rahmen der zweiten vorgeschlagenen Modifikation werden zusätzlich die imputierten Werte der nicht gezogenen Einheiten, die sich durch ein $\delta_i = 0$ mit $i \in M$ auszeichnen, nach Formel (4.3) adjustiert bzw. reimputiert. Die Erwartungswertberechnung bzw. die Wiederholung des ursprünglichen Imputationsschemas geschieht auf Basis der entsprechenden beobachtbaren Werte mit $\delta_i = 0$ und

$i \in R$. Der auf Grundlage der Adjustierung gewonnene Vektor wird für Modifikation 1 als $y_{M1,adjust}$ und für Modifikation 2 als $y_{M2,adjust}$ bezeichnet. Entsprechend resultieren bei Anwendung der Reimputation die Vektoren $y_{M1,reimp}$ und $y_{M2,reimp}$. Der Punktschätzer des Rescaling-Bootstrap wird in Abhängigkeit der Modifikation und ob die Reimputation oder Adjustierung zugrunde liegt, berechnet durch $\hat{\theta}_{b,resc1,adjust}^* = T(y_{M1,adjust}, w_{Resc}^*)$, $\hat{\theta}_{b,resc2,adjust}^* = T(y_{M2,adjust}, w_{Resc}^*)$, $\hat{\theta}_{b,resc1,reimp}^* = T(y_{M1,reimp}, w_{Resc}^*)$ bzw. $\hat{\theta}_{b,resc2,reimp}^* = T(y_{M2,reimp}, w_{Resc}^*)$.⁴ Die Varianzschätzung erfolgt wie beim Bootstrap üblich nach (4.21).

4.5.5 Rescaling-Bootstrap bei fehlenden Werten für mehrstufige Stichprobendesigns

Auch der in 4.5.3 beschriebene Rescaling-Bootstrap im Falle mehrstufiger Stichprobendesigns basiert auf dem Fall vollständiger Beobachtungen. Daher wird in diesem Abschnitt ein Ansatz entwickelt, wie das Verfahren auf den Fall fehlender Werte bzw. Imputation übertragen werden kann. Auch hier gilt es, die Varianz abzudecken, wie sie durch den ursprünglichen Ziehungs- und Imputationsprozess entstanden ist. Die Varianzen auf den einzelnen Stufen müssen richtig erfasst werden. Bezüglich der Berücksichtigung der Imputation beim Rescaling-Bootstrap müssen die im Rahmen des einstufigen Stichprobendesigns getätigten Aussagen auf das mehrstufige Design übertragen werden. Dabei ist die folgende Überlegung zu tätigen, wobei auch in diesem Abschnitt grundsätzlich auf ein zweistufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe abgestellt wird. Ausgehend von dem in Abschnitt 4.5.3 beschriebenen Verfahren ist es denkbar, dass eine bestimmte Einheit auf der letzten Stufe durch Ziehung ihrer PSU ($\delta_{hd} = 1$) auf der ersten Stufe oder über die zweite Stufe ($\delta_{hdi} = 1$) in die Substichprobe gelangt. Denkbar wäre es auch, dass sie auf beiden Stufe gezogen wird ($\delta_{hd} = 1$ oder $\delta_{hdi} = 1$) oder überhaupt nicht ($\delta_{hd} = 0$ oder $\delta_{hdi} = 0$). Dadurch ergeben sich vier Kategorien:

- K1: Eine Einheit gelangt nur über ihre PSU auf der ersten Stufe in die Substichprobe ($\delta_{hd} = 1$ und $\delta_{hdi} = 0$)
- K2: Eine Einheit gelangt nur auf der zweiten Stufe in die Substichprobe ($\delta_{hd} = 0$ und $\delta_{hdi} = 1$)
- K3: Eine Einheit gelangt auf beiden Stufen in die Substichprobe ($\delta_{hd} = 1$ und $\delta_{hdi} = 1$)
- K4: Eine Einheit gelangt auf keiner der Stufen in die Substichprobe ($\delta_{hd} = 0$ und $\delta_{hdi} = 0$)

Auch beim mehrstufigen Design gilt die Aussage des einstufigen Designs aus Abschnitt 4.5.4: Die Reimputation bzw. Adjustierung der imputierten Werte muss in jeder Substichprobe getrennt innerhalb der für die Substichprobe gezogenen wie auch der nicht gezogenen Elemente vollzogen werden. Werden die nicht für die Substichprobe gezogenen Imputationswerte außer Acht gelassen, geht je nachdem, ob eine Einheit ausgewählt worden ist oder nicht, einmal der reimputierte bzw. adjustierte Werte ein oder der Imputationswert der ursprünglichen Stichprobe. Dadurch kann über die Substichprobe hinweg zu viel Variation entstehen, welche auf diese Art im ursprünglichen Imputationsprozess nicht aufgetreten ist. Daher kann die Varianz überschätzt werden. Dies trifft jedoch nur

⁴ Für die detaillierte Berechnung von w_{Resc}^* siehe Abschnitt 4.5.2.

bei großen Auswahl­sätzen zu, wenn auch die nicht für die Substichprobe gezogenen Elemente ein größeres Gewicht besitzen. Eine Reimputation bzw. Adjustierung der imputierten Werte auf die gesamte Stichprobe würde wie im einstufigen Fall bei deterministischen Imputationsverfahren in jeder Substichprobe zu den originären Imputationswerten der ursprünglichen Stichproben führen. Dadurch würde über die Substichproben hinweg keine Variation entstehen und die Varianz wird unterschätzt.

Aber auch die Trennung in für die Substichprobe gezogenen und nicht gezogenen Einheiten, wie sie im einstufigen Fall verwendet worden ist, reicht hier nicht aus. Dafür existiert ein wichtiger Grund.

Unter denen am Anfang dieses Kapitels getätigten Annahmen der kleinen Auswahl­sätze oder des fixen Response-Vektors müssen nur die Stichprobenkomponente und bei den zufälligen Imputationsverfahren zusätzlich die Imputationskomponente der Varianz berücksichtigt werden. Des Weiteren wird der Fall betrachtet, dass fehlende Werte auf der letzten Stufe entstehen. Selbst in diesem Fall reicht eine Durchführung der Reimputation bzw. Adjustierung der imputierten Werte ausschließlich auf der letzten Stufe nicht aus. Wie in Abschnitt 4.1.3 beschrieben wurde, muss dem Umstand Rechnung getragen werden, dass Imputationen, selbst wenn sie nur auf Einheiten der letzten Stufe vollzogen worden sind, auch die Varianz übergeordneter Stufen beeinflussen. So gehen z.B. beim zweistufigen Design die imputierten Werte in die Berechnung der Totalwerte der einzelnen PSU ein und beeinflussen die Varianz dieser Stufe. Daher muss hier eine Reimputation bzw. Adjustierung der imputierten Werte in der Substichprobe sowohl auf der ersten als auch zweiten Stufe vorgenommen werden.

Aus diesen Gründen muss einerseits eine Trennung zwischen für die Substichprobe gezogenen und nicht gezogenen Elementen vollzogen werden und andererseits zwischen den verschiedenen Stufen. Und dies führt für ein zweistufiges Stichprobendesign zu den am Anfang dieses Abschnittes beschriebenen vier Kategorien. Die logische Konsequenz wäre damit, die Reimputation bzw. Adjustierung in jeder Substichprobe getrennt in allen vier Kategorien K1 bis K4 vorzunehmen. Jedoch gilt es zu beachten, dass die Einheiten in K2, wie in Abschnitt 4.5.3 beschrieben, beim Rescaling-Bootstrap den gleichen Gewichts­anpassungen unterliegen wie die auf keiner Stufe für die Substichprobe gezogenen Einheiten. Sie werden damit behandelt wie nicht für die Substichprobe gezogen. Um z.B. die Stichprobenkomponente korrekt zu erfassen, könnte es sinnvoll sein, dass die Reimputationen bzw. Adjustierungen der imputierten Werte mit dem Substichprobenbildungsprozess in Verbindung stehen. Die Konsequenz wäre die Zusammenlegung der Kategorien K2 und K4. Reimputationen bzw. Adjustierungen würden separat innerhalb der Kategorien K1, K3 sowie in dieser zusammengelegten Kategorie durchgeführt werden. Für eine Anwendung der Reimputation bzw. Adjustierungen separat in den Kategorien K1 bis K4 ohne Zusammenlegung könnte hingegen eine stabilere Varianz­schätzung aufgrund gleicher Gruppengrößen sprechen. Durch die Zusammenlegung entsteht eine größere und zwei kleinere Kategorien. Um dies zu überprüfen, werden beide in dieser Arbeit vorgeschlagenen Varianten zur Kategorienbildung gegenüber anderen Möglichkeiten der Kategorienabgrenzung zur Reimputation bzw. Adjustierung der imputierten Werte in der Simulationsstudie verglichen.

Zusammenfassend ergibt sich für die beiden Varianten das folgende allgemeine Vorgehen. Es werden Kategorien K_π gebildet, welche die zuvor beschriebene Trennung von für die Substichprobe gezogenen und nicht gezogenen Elementen und den verschiedenen Stufen berücksichtigen. Die Reimputation bzw. Adjustierung eines imputierten Wertes eines fehlenden Wertes in der Substichprobe innerhalb einer Kategorie $i \in M_{K_\pi}$ werden auf Basis der entsprechenden beobachtbaren Werte $i \in R_{K_\pi}$ der Kategorie vorgenommen. Dies

erfolgt unter Umständen unter Rückgriff auf Hilfsvariablen. Liegen Imputationsklassen vor, so erfolgen die Reimputationen und Adjustierungen der imputierten Werte innerhalb einer Kategorie zusätzlich innerhalb dieser Klassen. Es resultiert wie zuvor das reimputierte bzw. adjustierte Untersuchungsmerkmal mit dem Vektor $y_{K_\pi, reimp}$ bzw. $y_{K_\pi, adjust}$. Der Rescaling-Bootstrap-Punktschätzer wird berechnet durch $\hat{\theta}_{b, resc}^* = T(y_{K_\pi, reimp}, w_{Resc}^*)$ bzw. $\hat{\theta}_{b, resc}^* = T(y_{K_\pi, adjust}, w_{Resc}^*)$ und die Varianz geschätzt durch (4.21).⁵ Durch dieses Vorgehen ist es auch beim Rescaling-Bootstrap ohne Zurücklegen im Fall des mehrstufigen Stichprobendesigns möglich, fehlende Werte und Imputationen zu berücksichtigen.

4.6 Balanced-Repeated-Replication

4.6.1 Allgemeine Beschreibung

Die letzte in dieser Arbeit verwendete Resampling-Methode ist Balanced-Repeated-Replication (BRR). Die Anfänge dieses Verfahrens gehen zurück auf MCCARTHY (1966) bzw. MCCARTHY (1969). Für den BRR sind geschichtete Stichprobendesigns Anwendungsvoraussetzung. In ihrer Basisversion geht die Methode von zwei Einheiten pro Schicht aus. Für die Substichprobe wird eine dieser Einheiten gezogen. Es wird von einem *Half-Sample* gesprochen. Würden die Substichproben über alle möglichen Kombinationen von Half-Samples gebildet werden, gilt $\Omega = 2^H$. Bei einer großen Anzahl von Schichten würde dies zu einem erheblichen Aufwand führen. Um die Komplexität zu reduzieren, kommt es zu einer *ausbalancierten* Auswahl von Ω Half-Samples mit $\Omega \ll 2^H$ (vgl. SHAO et al., 1998, S. 822, MÜNNICH, 2008, S. 325 f., WOLTER, 2007, S. 107 ff., DAVISON und SARDY, 2004a, S. 18, BRUCH et al., 2011, S. 19).

Wird für diese Auswahl eine $\Omega \times H$ Matrix benutzt, mit dem (a, h) -ten Eintrag $\delta_{ah} = +1$ oder -1 , in Abhängigkeit davon ob das erste oder das zweite Element in der jeweiligen Schicht ausgewählt wird, dann wird von einer *ausbalancierten* Auswahl gesprochen, wenn gilt:

$$\sum_{a=1}^{\Omega} \delta_{ah} \delta_{ah'} = 0 \quad \forall h \neq h', \quad (4.30)$$

(vgl. SHAO et al., 1998, S. 822, MÜNNICH, 2008, S. 325 f., WOLTER, 2007, S. 107 ff., DAVISON und SARDY, 2004a, S. 18, BRUCH et al., 2011, S. 19).

Dies wird beim BRR häufig mittels einer Hadamard-Matrix realisiert. Deren Zeilen stehen für das jeweilige Half-Sample und die Spalten für die jeweilige Schicht. Die Anzahl der Replikationen ist festgelegt durch $H + 1 \leq \Omega \leq H + 4$. Die Zeilen und Spalten einer Hadamard-Matrix sind orthogonal. Damit werden die Half-Sample unabhängig realisiert und die Erfüllung der Bedingung (4.30) ist gegeben (vgl. DAVISON und SARDY, 2004a, S. 18, MÜNNICH, 2008, S. 325 f., SHAO et al., 1998, S. 822 und BRUCH et al., 2011, S. 19). Da genau die Hälfte der Einheiten für die Substichprobe gezogen wird, müssen die Designgewichte zur Berechnung von $\hat{\theta}_a$ mit zwei multipliziert werden (vgl. DAVISON und SARDY, 2004a, S. 17).

In der Praxis bestehen Schichten selten nur aus zwei Einheiten. Ist die Anzahl der Elemente größer als zwei, werden die Einheiten einer Schicht zufällig in zwei Gruppen eingeteilt. Diese besitzen die Größe $n_h^* = \lfloor n_h/2 \rfloor$ und $n_h - n_h^*$. Für das Half-Sample wird anstelle

⁵ Für die detaillierte Berechnung von w_{Resc}^* siehe Abschnitt 4.5.3.

eines Elementes eine dieser Gruppen nach dem zuvor beschriebenen Vorgehen gezogen (vgl. WOLTER, 2007, S. 128 ff., DAVISON und SARDY, 2004a, S. 17 ff. RAO und SHAO, 1996, S. 343 f., RAO und SHAO, 1996, MÜNNICH, 2008, S. 326, SHAO et al., 1998, S. 824). Auch in diesem Fall muss eine Anpassung der Designgewichte erfolgen, welche zudem auch der Tatsache Rechnung tragen müssen, dass Stichproben in der Realität häufig ohne Zurücklegen gezogen werden. Damit gilt es auch hier, die Endlichkeitskorrektur zu berücksichtigen. Ein solcher Ansatz zur Gewichtsangpassung beim BRR basierend auf dieser Gruppenstruktur und Berücksichtigung der Endlichkeitskorrektur wird in DAVISON und SARDY (2004a, S. 18 f.) gegeben. Demnach werden die Gewichte angepasst durch:

$$w_{h,i}^* := \begin{cases} w_{hi} \cdot \left[1 + \left\{ \frac{(n_h - n_h^*) \cdot (1 - f_h)}{n_h^*} \right\}^{1/2} \right], & \delta_{ah} = 1, \\ w_{hi} \cdot \left[1 - \left\{ \frac{n_h^* \cdot (1 - f_h)}{n_h - n_h^*} \right\}^{1/2} \right], & \delta_{ah} = -1. \end{cases} \quad (4.31)$$

In (4.31) ist zu sehen, dass eine Einbeziehung der Endlichkeitskorrektur $(1 - f_h)$ in die Gewichtsangpassung erfolgt. Damit wird dem Ziehungsmodell ohne Zurücklegen Rechnung getragen. Die angepassten Gewichte der einzelnen Einheiten werden zu dem Gewichtevektor w_{BRR} zusammengefasst. Dieses Vorgehen umfasst wieder alle Einheiten der Stichprobe, wobei die für die Substichprobe gezogenen Elemente nach (4.31) zumindest bei kleineren Auswahlsätzen ein deutlich größeres Gewicht besitzen. Die nicht gezogenen Elemente gewinnen aber auch hier mit größerem Auswahlsatz immer mehr an Bedeutung. Zur Berechnung des Punktschätzers in der Substichprobe wird daher der Vektor der Beobachtungsvariable in Höhe der Stichprobengröße y_S verwendet. Es gilt:

$$\hat{\theta}_{BRR,a} = T(y_S, w_{BRR}). \quad (4.32)$$

Für den Varianzschätzer gilt unter Verwendung von (4.2):

$$\hat{V}_{BRR}(\hat{\theta}) = \frac{1}{\Omega} \cdot \sum_{a=1}^{\Omega} \left(\hat{\theta}_{BRR,a} - \bar{\theta}_{BRR} \right)^2, \quad (4.33)$$

mit $\bar{\theta}_{BRR} = 1/\Omega \cdot \sum_{r=1}^{\Omega} \hat{\theta}_{BRR,a}$.

Dieses Vorgehen kann zu einem sehr variablen BRR-Varianzschätzer führen. Dies gilt insbesondere für eine kleine Schichtanzahl. Daraus resultiert eine geringe Anzahl an Replikationen. Um dieser Variabilität zu begegnen, besteht eine Möglichkeit darin, in jeder Replikation die Gruppierung der Einheiten Ψ mal zu wiederholen und den Varianzschätzer Ψ Mal zu berechnen. Anschließend erfolgt die Durchschnittsbildung über die einzelnen Varianzschätzer einer Replikation (vgl. RAO und SHAO, 1996, S. 344).

4.6.2 Vergleich zwischen BRR und Rescaling-Bootstrap

Zwischen den Methoden, welche unter der Kategorie Bootstrap fallen, und dem BRR existieren normalerweise große Unterschiede, insbesondere bezüglich der Substichprobenziehung. Jedoch zeigen sich hinsichtlich des Rescaling-Bootstrap ohne Zurücklegen und

des BRR in der Variante von DAVISON und SARDY (2004a) einige Gemeinsamkeiten. Diese sind herauszustellen, da sie bezüglich der Anwendung des BRR in dieser Arbeit unter fehlenden Werten und Imputation von Bedeutung sind.

So wird bei beiden Methoden als Substichprobe die Hälfte der ursprünglichen Stichprobengröße gezogen, wenn diese gerade sein sollte. Das Ziehungsmodell ist bei beiden Verfahren ohne Zurücklegen und zur Integration der Endlichkeitskorrektur werden alle Einheiten der ursprünglichen Stichprobe für die Berechnung des Punktschätzers berücksichtigt. Die Gewichtsadjustierungen nach (4.25) und (4.31) sind im Falle eines einstufigen stratifizierten Designs identisch.

Wird z.B. beim Rescaling-Bootstrap eine Einheit nicht gezogen und damit $\delta_{hi} = 0$, entfällt der letzte Term in (4.25) und bei Einsetzen von λ_h resultiert:

$$w_{hi}^* = w_{hi} \left(\cdot 1 - \sqrt{(1 - f_h) \cdot \frac{n_h^*}{(n_h - n_h^*)}} \right). \quad (4.34)$$

Dies entspricht der Gewichtsadjustierung für Einheiten des BRR von DAVISON und SARDY (2004a) in (4.31), welche nicht für das Half-Sample gezogen worden sind. Wird hingegen eine Einheit ausgewählt ($\delta_{hi} = 1$) ergibt sich bei Einsetzen von λ_h in (4.25):

$$\begin{aligned} w_{hi}^* &= w_{hi} \cdot \left[1 - \sqrt{(1 - f_h) \cdot \frac{n_h^*}{(n_h - n_h^*)}} + \frac{n_h}{n_h^*} \cdot \sqrt{(1 - f_h) \cdot \frac{n_h^*}{(n_h - n_h^*)}} \right] \\ &= w_{hi} \cdot \left[1 - \left(\sqrt{(1 - f_h) \cdot \frac{n_h^*}{(n_h - n_h^*)}} \cdot \left(1 - \frac{n_h}{n_h^*} \right) \right) \right] \\ &= w_{hi} \cdot \left[1 - \frac{n_h^* - n_h}{n_h^*} \cdot \sqrt{(1 - f_h) \cdot \frac{n_h^*}{(n_h - n_h^*)}} \right] \\ &= w_{hi} \cdot \left[1 + \frac{n_h - n_h^*}{n_h^*} \cdot \sqrt{(1 - f_h) \cdot \frac{n_h^*}{(n_h - n_h^*)}} \right] \\ &= w_{hi} \cdot \left[1 + \sqrt{(1 - f_h) \cdot \frac{n_h - n_h^*}{n_h^*}} \right]. \end{aligned}$$

Die letzte Zeile entspricht damit der Gewichtsadjustierung derjenigen Elemente des BRR in (4.31), welche für das Half-Sample gezogen worden sind.

Wie aber auch erwähnt, gibt es wichtige Unterschiede zwischen beiden Methoden. Beim Rescaling-Bootstrap ohne Zurücklegen erfolgt die Auswahl der Elemente mittels einer Zufallsauswahl. Es wird eine individuell festgelegte Anzahl von B Substichproben gezogen und die Varianz wie beim Bootstrap üblich auf Basis der Monte-Carlo-Approximation berechnet. Beim BRR hingegen werden Gruppen gebildet und z.B. mittels Hadarmard-Matrizen ausgewählt. Ziel ist die Gewährleistung einer ausbalancierten Auswahl. Dadurch können gerade Ausreißer bei ihrer Ziehung für die Substichprobe nicht so stark ins Gewicht fallen wie bei der einfachen Zufallsauswahl. Die Zahl der Replikationen kann ohne wiederholte Gruppierung nicht individuell festgelegt werden, sondern ist durch die Anzahl der Schichten determiniert. Auch sprechen beide Verfahren in ihrer grundsätzlichen Ausrichtung andere Stichprobendesigns an. Der Rescaling-Bootstrap ohne Zurücklegen wurde u.a. speziell für mehrstufige Stichprobendesigns entwickelt. Aus den in Abschnitt 4.5.3 genannten Gründen entfaltet er gerade hier seine Vorteile. Er kann aber natürlich auch im Falle von einstufigen stratifizierten Designs angewendet werden. Der BRR stellt besonders auf diese Designs ab. Seine Übertragung auf mehrstufige Stichprobendesigns

kann gemäß des in Abschnitt 4.1.3 erläuterten zweiten Vorgehens vorgenommen werden. Jedoch ist diese Anwendung beim mehrstufigen Design deutlich komplizierter und rechenintensiver wie für den Rescaling-Bootstrap ohne Zurücklegen. Zur Erfassung der Varianz aller Stufen muss der BRR auf allen Stufen separat angewendet werden. Wird zum Beispiel ein stratifiziertes zweistufiges Design betrachtet, so werden auf der ersten Stufe die PSU in den einzelnen Schichten zu Gruppen zusammengefasst. Diese Gruppierung von PSU kann unter Umständen negative Effekte mit sich bringen, bei sehr homogenen PSU. Auf der zweiten Stufen können die PSU als Schichten behandelt werden. Dadurch ist die Durchführung des BRR möglich. Durch diese Anwendung des BRR auf allen Stufen kann seine Berechnung sehr aufwendig werden, gerade wenn das Stichprobendesign viele Stufen beinhaltet. Beim Rescaling-Bootstrap ohne Zurücklegen erfolgt eine andere Vorgehensweise. Hier werden alle Stufen des Designs über die Gewichtung innerhalb der Berechnung des Punktschätzers integriert, wodurch der Varianzschätzer nicht separat auf beiden Stufen berechnet werden muss.

4.6.3 BRR bei fehlenden Werten und Imputation

Die grundlegende Behandlung des BRR unter fehlenden Werten und Imputation wird in SHAO et al. (1998) gezeigt. Dort wird jedoch von einer vernachlässigbaren Endlichkeitskorrektur ausgegangen. Wird stattdessen zur Integration dieser Korrektur der Ansatz über (4.31) gewählt, muss auf anderem Wege vorgegangen werden. Hier stellt sich wieder das bekannte Problem, dass durch die Gewichts Anpassung in (4.31) auch die Einheiten in die Berechnung des Punktschätzers mit eingehen, welche für die Substichprobe, in diesem Fall das Half-Sample, nicht gezogen wurden. Diese Einheiten bekommen mit größeren Auswahlätzen ein immer größeres Gewicht, da wie in Abschnitt 4.6.2 beschrieben, die Gewichts Anpassung im einstufigen stratifizierten Design derjenigen des Rescaling-Bootstrap ohne Zurücklegen entspricht. Es ergibt sich damit die gleiche Problematik wie bei dieser Methode. Daher wird im einstufigen Design zur Berücksichtigung von fehlenden Werten und Imputation vorgeschlagen, auch hier die in 4.5.4 beschriebene zweite Modifikation anzuwenden.

Damit wird eine separate Reimputation bzw. Adjustierung der imputierten Werte zwischen den für die Substichprobe gezogenen ($\delta_{ah} = 1$) bzw. nicht gezogenen Elemente ($\delta_{ah} = -1$) vorgenommen. Da auch hier die Substichprobengröße der Hälfte der ursprünglichen Stichprobengröße entspricht, gelten die Erläuterungen des Rescaling-Bootstrap ohne Zurücklegen (vgl. hierzu auch SHAO, 2002). Bei deterministischen Imputationsverfahren können sowohl die Reimputation wie auch die Adjustierung angewendet werden, bei zufälligen Imputationsverfahren nur die Adjustierung. In Änderung zu (4.32) gilt damit:

$$\hat{\theta}_{BRR,a} = T(y_{M2,adjust}, w_{BRR}) \quad (4.35)$$

bzw.

$$\hat{\theta}_{BRR,a} = T(y_{M2,reimp}, w_{BRR}). \quad (4.36)$$

Es wird nur die zweite Modifikation berücksichtigt. Der Grund hierfür stellen die in Abschnitt 4.5.4 erläuterten Nachteile der Modifikation 1 dar. Die über Modifikation 2 adjustierten bzw. reimputierten Vektoren des Untersuchungsmerkmals werden durch $y_{M2,adjust}$ respektive $y_{M2,reimp}$ gekennzeichnet. Die Varianzschätzung erfolgt weiterhin über Formel (4.33).

Im mehrstufigen Stichprobendesign existieren die in Abschnitt 4.6.2 beschriebenen Unterschiede zwischen Rescaling-Bootstrap ohne Zurücklegen und dem BRR. Letztere Methode wird im Rahmen dieser Arbeit beim mehrstufigen Design gemäß (4.7) bzw. dem in Abschnitt 4.6.2 beschriebenen Anwendung durchgeführt. Es muss daher das Vorgehen des Rescaling-Bootstrap in 4.5.5 auf die separate Anwendung des BRR auf den verschiedenen Stufen übertragen werden. Auch hier gilt es zu beachten, dass sich die Imputation niedriger Stufen auf die Varianz übergeordneter Stufen auswirkt. Damit muss die Reimputation bzw. Adjustierung bei der separaten Anwendung des BRR auf den einzelnen Stufen immer mit durchgeführt werden. Auf jeder Stufe muss auch hier eine Trennung von für die Substichprobe gezogen und nicht gezogenen Elementen vollzogen werden. Wird von einem zweistufigen Design ausgegangen, resultieren auf der ersten Stufe der Punktschätzer $\hat{\theta}_{BRR,a} = T(y_{M2,adjust,1}, w_{BRR,1})$ bzw. $\hat{\theta}_{BRR,a} = T(y_{M2,reimp,1}, w_{BRR,1})$. $y_{M2,adjust,1}$ bzw. $y_{M2,reimp,1}$ stellt die reimputierte bzw. adjustierten Vektor des Untersuchungsmerkmals dar. Die Reimputation bzw. Adjustierung wird getrennt für fehlende Werte auf der letzten Stufe durchgeführt, je nachdem ob ihre PSU für die Substichprobe gezogen wurde oder nicht. Der Gewichtevektor $w_{BRR,1}$ beinhaltet die bezüglich der Einheiten der ersten Stufe nach (4.31) angepassten Designgewichte der Einheiten der letzten Stufe $w_{i,1}^* = w_{hd}^* \cdot w_{hdi}$. Es gilt:

$$w_{hd}^* := \begin{cases} w_{hd} \cdot \left[1 + \left\{ \frac{(l_h - l_h^*) \cdot (1 - f_h)}{l_h^*} \right\}^{1/2} \right], & \delta_{ahd} = 1, \\ w_{hd} \cdot \left[1 - \left\{ \frac{l_h^* \cdot (1 - f_h)}{l_h - l_h^*} \right\}^{1/2} \right], & \delta_{ahd} = -1. \end{cases} \quad (4.37)$$

Auf der zweiten Stufe wird hingegen mittels des BRR auf die Punktschätzer $\hat{\theta}_{BRR,o} = T(y_{M2,adjust,2}, w_{BRR,2})$ bzw. $T(y_{M2,reimp,2}, w_{BRR,2})$ abgestellt. $y_{M2,adjust,2}$ bzw. $y_{M2,reimp,2}$ stellen auch hier den reimputierten bzw. adjustierten Vektor des Beobachtungsmerkmals dar. Die Reimputation bzw. Adjustierung findet hier jedoch separat zwischen Imputationswerten statt, die auf der letzten Stufe für die Substichprobe gezogen bzw. nicht gezogen wurden. $w_{BRR,2}$ ist dementsprechend der Vektor der Gewichte $w_{hdi,2}^*$, welche berechnet werden durch:

$$w_{hdi,2}^* := \begin{cases} w_{hdi} \cdot \left[1 + \left\{ \frac{(n_{hd} - n_{hd}^*) \cdot (1 - f_{hd})}{n_{hd}^*} \right\}^{1/2} \right], & \delta_{ahdi} = 1, \\ w_{hdi} \cdot \left[1 - \left\{ \frac{n_{hd}^* \cdot (1 - f_{hd})}{n_{hd} - n_{hd}^*} \right\}^{1/2} \right], & \delta_{ahdi} = -1. \end{cases} \quad (4.38)$$

n_{hd} stellt weiterhin die Anzahl der Elemente in der PSU d dar und des Weiteren gilt $n_{hd}^* = \lfloor n_{hd}/2 \rfloor$.

Wird des weiterhin davon ausgegangen, dass L_h/l_h über die Schichten konstant ist, z.B. im Falle einer proportionalen Allokation oder einer gleichen PSU-Größe in der Grundgesamtheit und Stichprobe auf der ersten Stufe, kann die Varianz unter Rückgriff auf (4.7) geschätzt werden durch:

$$\begin{aligned} \hat{V}_{mult}(\hat{\theta}_{2St}) &= \sum_{a=1}^{\Omega_1} \left(\hat{\theta}_{BRR,a} - \frac{1}{\Omega_1} \cdot \sum_{v=1}^{\Omega_1} \hat{\theta}_{BRR,v} \right)^2 \\ &+ \frac{L_h}{l_h} \cdot \sum_{o=1}^{\Omega_2} \left(\hat{\theta}_{BRR,o} - \frac{1}{\Omega_2} \sum_{j=1}^{\Omega_2} \hat{\theta}_{BRR,j} \right)^2. \end{aligned} \quad (4.39)$$

Die Anwendung von Schätzer (4.6) ist nicht zu empfehlen. Dies würde bedeuten, dass der BRR auf der zweiten Stufe innerhalb einer PSU angewendet werden muss. Ohne Schichtung innerhalb der PSU entspreche dies einer Anwendung des BRR, welche dem Fall des Vorliegens einer Schicht gleichkäme. In diesem Fall wäre der Schätzer sehr variabel. Aber auch das Vorgehen aus (4.39) kann Probleme mit sich bringen. Die Problematik der Gruppierung sehr homogener PSU wurde schon in Abschnitt 4.6.2 beschrieben. Des Weiteren kann dieses Vorgehen nicht mehr im Falle eines über die Schichten variierenden Terms L_h/l_h angewendet werden.

4.7 Resampling-Verfahren unter Berücksichtigung der Varianzkomponente bezüglich fehlender Werte

4.7.1 Allgemeine Problematik

Die bisher beschriebenen Resampling-Verfahren gehen von einer vernachlässigbaren Varianzkomponente bezüglich fehlender Werte der Varianz aus. Damit ist die zweite Komponente V_{NR}^1 bzw. \tilde{V}_{NR}^1 in (3.23) respektive (3.26) gemeint. Diese Komponente kann in Fällen außer Acht gelassen werden, wenn entweder die fehlenden Werte in der Grundgesamtheit nur einmalig erzeugt werden oder der Gesamtauswahlsatz klein ist. Liegt jedoch ein großer Gesamtauswahlsatz vor und sind fehlende Werte in der Grundgesamtheit bzw. der Stichprobe variabel, führen die beschriebenen Methoden zu einer deutlichen Unterschätzung der Varianz. Der Grund liegt darin, dass die Resampling-Verfahren über die Substichproben hinweg mit dem gegebenen Response-Vektor arbeiten. Jede Einheit ist als beobachtet oder fehlend eindeutig festgelegt. Der Response-Vektor wird vor der Reimputation bzw. Adjustierung in jeder Substichprobe nicht neu erzeugt. Damit kann die Variation, welche durch die wiederholte Erzeugung des Response-Vektors in jedem Simulationsdurchlauf resultiert, entweder in der Grundgesamtheit oder der Stichprobe, nicht berücksichtigt werden.

Dieser Sachverhalt lässt sich am besten für den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit zeigen (vgl. hierzu und im Folgenden auch die Erläuterungen in MASHREGHI et al., 2014 bzw. HAZIZA, 2010). Der Grund liegt darin, dass Resampling-Verfahren im Falle des Vorliegens fehlender Werte und deren Kompensation über Imputation eher bezüglich dieses Vorgehens analysiert werden können gegenüber dem Auftreten fehlender Werte in der Stichprobe. Dies betrifft insbesondere die Ausrichtung des äußeren Vorganges auf den Ausfallmechanismus und den inneren Vorgang auf das Stichprobendesign. Die Grundlage bilden die beiden Formeln aus (3.23) bzw. (3.26). Werden zunächst die deterministischen Imputationsverfahren herangezogen, entfällt die Komponente \tilde{V}_{Imp}^1 und es werden die beiden Komponenten $V_S^1 = E_q V_p(\hat{\theta}^I | z)$ und $V_{NR}^1 = V_q E_p(\hat{\theta}^I | z)$ aus (3.23) betrachtet. Nach MASHREGHI et al. (2014, S. 146) muss ein valider

Schätzer für V_S^1 die Stichprobenvariabilität des imputierten Schätzers konditioniert auf den Response-Vektor z erfassen. Und genau diesen Vorgang decken Resampling-Verfahren ab. Sie reproduzieren für einen gegebenen Response-Vektor z über die Selektion von Substichproben die beschriebene Variation des Schätzers durch die Stichprobenziehung. Wie anfangs erläutert geschieht dies nur für einen konkreten Response-Vektor. Die zweite Komponente $V_{NR}^1 = V_q E_p(\hat{\theta}^I | z)$, welche die Variation über die verschiedenen Response-Vektoren beschreibt, wird dadurch nicht abgedeckt (vgl. MASHREGHI et al., 2014, S. 146). Diese Analyse kann für den Fall der Entstehung fehlender Werte in der Grundgesamtheit leicht vollzogen werden, da auch für Resampling-Verfahren die einmalige Generierung des Response-Vektors zeitlich vor der Substichprobenziehung liegt. Für den Fall der Entstehung fehlender Werte in der Stichprobe lässt sich die Abdeckung des Anteils an der Gesamtvarianz der Resampling-Verfahren nicht so einfach demonstrieren. Denn hier sind die einzelnen Varianzkomponenten anders definiert. Nach (3.24) stellt der äußere Vorgang die Stichprobenziehung dar und der innere Vorgang bezieht sich auf den Antwortmechanismus. Damit erfolgt hier die Stichprobenziehung vor der Generierung des Response-Vektors. Wie beschrieben, ist es bei Resampling-Verfahren gegenteilig zu betrachten. Die Simulationsstudie wird später zeigen, dass die erste Komponente \tilde{V}_S^2 bei dieser Zerlegung nicht mit dem durch die Resampling-Methoden abgedeckten Anteil übereinstimmt. Nach der Varianzzerlegung in (3.26) sind für zufällige Imputationsverfahren drei Terme zu berücksichtigen. Zusätzlich kommt die Imputationskomponente \tilde{V}_{Imp}^1 hinzu. Der zweite Term aus (3.26) $\tilde{V}_{NR}^1 = V_q E_p E_I(\hat{\theta}^I | s, z)$, die Varianzkomponente durch fehlende Werte, wird durch die klassischen Resampling-Verfahren auch hier nicht erfasst. Nach MASHREGHI et al. (2014, S. 146) werden die erste und dritte Komponente $\tilde{V}_S^1 = E_q V_p E_I(\hat{\theta}^I | s, z)$ bzw. $\tilde{V}_{Imp}^1 = E_q E_p V_I(\hat{\theta}^I | s, z)$ für kleine Auswahlsätze erfasst. Bei großen Auswahlsätzen zeigen sich aber auch hier Probleme. So liegt nach den Autoren für bestimmte Resampling-Verfahren eine schlechte Approximationsgüte der gemeinsamen Verteilung aus Stichproben und Donoren vor. Dies gilt für Resampling-Methoden, welche Endlichkeitskorrekturen beim Modell ohne Zurücklegen über Reskalierungen berücksichtigen (vgl. MASHREGHI et al., 2014, S. 146). Damit ist z.B. der Rescaling-Bootstrap aus Abschnitt 4.5.2 und der BRR aus Abschnitt 4.6 betroffen. Auch hier können Unterschätzungen der Varianz die Folge sein.

So wird nach MASHREGHI et al. (2014) zum Beispiel bei der ursprünglichen Version des Rescaling-Bootstrap aus 4.5.1 bei (gewichteter) Hot-Deck-Random-Imputation im Falle der einfachen Zufallsstichprobe und uniformer Ausfallmechanismus die Varianz $\tilde{V}_S^1 + (1 - f) \cdot \tilde{V}_{Imp}^1$ abgedeckt. Der Rescaling-Bootstrap ohne Zurücklegen aus Abschnitt 4.5.2 deckt die gleiche Quantität ab. Dies werden die Simulationsergebnisse aus Abschnitt 5.3.5 zeigen. Der BRR aus 4.31 verwendet, wie in Abschnitt 4.6.2 beschrieben, die gleichen Reskalierungen wie der Rescaling-Bootstrap ohne Zurücklegen, welche für die schlechte Approximationsgüte verantwortlich sind. Nicht erfasst werden in diesem Fall die Komponenten $\tilde{V}_{NR}^1 + f \cdot \tilde{V}_{Imp}^1$. Damit führen die Verfahren bei einem kleinen Auswahlsatz f zu unverzerrten Schätzern, jedoch auch hier bei einem großen Auswahlsatz zu Unterschätzungen der Varianz (vgl. MASHREGHI et al., 2014, S. 146).

Und genau dieser letzte Punkt muss noch einmal herausgestellt werden. Die bisher erläuterten Resampling-Verfahren können bei kleinen Auswahlsätzen ohne Probleme angewendet werden, um die Varianz bei Auftreten von fehlenden Werten und der Durchführung von Imputationen zu schätzen. In diesem Fall ist wie in 3.4.3 erläutert die Varianzkomponente durch fehlende Werte V_{NR}^1 bzw. \tilde{V}_{NR}^1 vernachlässigbar klein. Liegen jedoch große Auswahlsätze vor, insbesondere ein großer Gesamtauswahlsatz f , müssen die Resampling-Verfahren angepasst werden. Und hier werden in HAZIZA (2010) und MASHREGHI et al.

(2014) verschiedene Ansätze gegeben, welche im nächsten Abschnitt näher erläutert werden.

4.7.2 Verfahren zur Berücksichtigung der Varianzkomponente bezüglich fehlender Werte

Alle diese Ansätze gehen von der strengen Annahme des uniformen Antwortmechanismus aus. Die Anwendung von Imputationsklassen ist daher sinnvoll. Außerdem wurden sie für Fälle hergeleitet, bei welchen die Imputationsklassen den Schichten entsprechen. Die Methoden in den Arbeiten von MASHREGHI et al. (2014) wurden insbesondere an der ursprünglichen Version des Rescaling-Bootstrap nach RAO und WU (1988) demonstriert. Die Probleme dieser Methode wurden in 4.5.1 beschrieben.

Die intuitiv naheliegendste Methode ist es, die Erzeugung der fehlenden Werte innerhalb der Substichproben zu reproduzieren (vgl. im Folgenden MASHREGHI et al., 2014). Die Response-Vektoren z_{hi}^* werden in jeder Stichprobe neu erzeugt und anschließend imputiert. MASHREGHI et al. (2014) schlagen vor, die Response-Vektoren über eine Bernoulli Verteilung zu erzeugen mit:

$$z_{hi}^* \sim \text{Bernoulli}(\hat{p}_{Rh}). \quad (4.40)$$

Falls die Schichten den Imputationsklassen entsprechen gilt:

$$\hat{p}_{Rh} = r_h/n_h. \quad (4.41)$$

Ansonsten müssten die Imputationsklassen oder sonstige Abgrenzungen zur Gewährleistung des uniformen Antwortmechanismus zugrunde gelegt werden.

Die Erzeugung der fehlenden Werte und die Ziehung der Einheiten für die Substichprobe muss unabhängig erfolgen. Werden bei dem Verfahren Reskalierungen zur Berücksichtigung der Endlichkeitskorrektur durchgeführt, muss es zur Anpassung dieser Faktoren kommen. Und dadurch kann dieses Vorgehen sehr aufwendig werden. Diese Anpassung unterscheidet sich nicht nur zwischen den Imputationsmethoden, sondern auch zwischen den Resampling-Verfahren. So wird z.B. bei der Mittelwertimputation beim Rescaling-Bootstrap nach RAO und WU (1988) statt $C_h = ((1 - f_h) \cdot n_h^*/(n_h - 1))$ aus Abschnitt 4.5.1 $C_h^I = ((1 - \hat{p}_{Rh} \cdot f_h) \cdot n_h^*/(r_h - 1))$ verwendet.⁶ Wenn für $\hat{p}_{Rh} \cdot f_h = r_h/N_h$ gilt, sind sich die Faktoren sehr ähnlich. Statt der Stichprobengröße wird die Anzahl der antwortenden Einheiten genommen (vgl. MASHREGHI et al., 2014, S. 146 ff.).

Wie sich an der Beschreibung dieses Verfahrens erkennen lässt, ist die Umsetzung sehr aufwendig und computerintensiv. Wie schon beschrieben, muss die Reskalierungskonstante für jede Imputationsmethode und Resampling-Methode separat hergeleitet werden. In jeder Substichprobe wird zudem nicht nur die Ziehung und die Imputation wiederholt, sondern auch die Generierung des Response-Vektors. Das kann je nach Vorgehen zu deutlich längeren Rechenzeiten führen. Im Abschnitt 4.3 ihres Artikels geben MASHREGHI et al. (2014) zudem an, wie dieses Vorgehen auch auf Methoden übertragen werden kann, bei denen die Reskalierung auf die Gewichte und weniger auf die Untersuchungsvariable vorgenommen wird.

⁶ Die Rescaling-Faktoren für weitere Imputationsmethoden werden in Tabelle 1 in Abschnitt 4.2 in MASHREGHI et al. (2014) gegeben.

Eine weitere Möglichkeit, die Varianzkomponente bezüglich fehlender Werte zu berücksichtigen, besteht darin, den direkten Schätzer für V_{NR}^1 einfach zum Varianzschätzer der Resampling-Methode hinzu zu addieren (vgl. HAZIZA, 2010, S. 8). Bei deterministischen Imputationsmethoden decken die Verfahren die erste Komponente der Varianz ab. Durch die Addition des direkten Schätzers der zweiten Komponente könnte die gesamte Varianz erfasst werden. So würde bei Mittelwertimputation bei der einfachen Zufallsstichprobe dem Resampling-Varianzschätzer der in (3.45) beschriebenen Schätzer $\hat{V}_{NR}^1 = m/N \cdot s_R^2/r$ hinzugefügt werden. Voraussetzung für dieses Vorgehen ist, dass ein direkter Varianzschätzer der zweiten Stufe vorliegt. Außerdem müssen die Bedingungen aus Abschnitt 3.6 zur Herleitung des Schätzers beachtet werden. Der direkte Schätzer der zweiten Stufe müsste zudem immer separat mitberechnet werden, was unter Umständen z.B. bei mehrstufigen Stichprobendesigns sehr aufwendig werden kann.

Statt die zweite Varianzkomponente separat zu schätzen und additiv zu berücksichtigen, besteht nach MASHREGHI et al. (2014, S. 150) eine weitere Möglichkeit darin, die Varianzkomponente bezüglich fehlender Werte durch Verwendung von Korrekturfaktoren bezüglich der Resampling-Methode zu berücksichtigen. Dies wird nachfolgend ausführlich erläutert.

4.7.3 Berücksichtigung der Varianzkomponente bezüglich fehlender Werte über Korrekturfaktoren

4.7.3.1 Allgemeines Vorgehen

Bei deterministischen Imputationsverfahren ist der Ausgangspunkt wiederum die Überlegung, dass Resampling-Verfahren die Stichprobenkomponente V_S^1 abdecken. Ziel ist es einen bestimmten Korrekturfaktor α_I zu finden, welcher mit der Varianzschätzung der Resampling-Methode multipliziert wird, um die gesamte Varianz V_{ges} abzudecken. Damit muss gelten:

$$V_{ges} = \alpha_I \cdot V_S^1. \quad (4.42)$$

Dieser Korrekturfaktor α_I muss letztendlich geschätzt werden. Das Vorgehen von MASHREGHI et al. (2014) wird im Rahmen dieser Arbeit in vier Schritten zusammengefasst. Demnach ergibt sich der folgende allgemeine Ablauf:

1. *Varianzermittlung*: Die Varianz beider Komponenten wird zunächst approximativ hergeleitet. Hierzu ist es meist notwendig, bestimmte Annahmen zu treffen bzw. von starken Vereinfachungen auszugehen. MASHREGHI et al. (2014) gehen in ihrem Fall von der einfachen Zufallsstichprobe, von einem uniformen Antwortmechanismus sowie der Entstehung variabler fehlender Werte in der Grundgesamtheit aus. Korrekturfaktoren für den Fall eines fixen Response-Vektors sind nur für zufällige Imputationsverfahren notwendig. Ein Korrekturfaktor für Hot-Deck-Random-Imputation für diesen Fall wird im Rahmen dieser Arbeit nach Vorstellung des allgemeinen Ablaufs und zwei konkreten Beispielen der Korrekturfaktoren für den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit angegeben und im Anhang hergeleitet. Für den Fall der Entstehung fehlender Werte in der Stichprobe ist die Herleitung nicht mehr so einfach möglich. Der Grund liegt darin, dass sich die beiden Komponenten vom Fall der Entstehung fehlender Werte in der Grundgesamtheit in ihrer Höhe unterscheiden. Resampling-Verfahren decken hier nicht die Komponente

V_S^2 ab. Hier wird im Rahmen der Simulationsstudie geschaut, inwieweit die Korrekturfaktoren aus dem Fall variabler fehlender Werte übernommen werden können.

2. *Herleitung des Korrekturfaktors*: Bei deterministischen Imputationsverfahren gilt:

$$V_{ges} = V_S^1 + V_{NR}^1. \quad (4.43)$$

Damit kann (4.42) umgeformt werden zu:

$$\alpha_I = \frac{V_S^1 + V_{NR}^1}{V_S^1}. \quad (4.44)$$

Es wird hier auf die theoretisch hergeleiteten Varianzen V_S^1 und V_{NR}^1 aus Schritt 1 zurückgegriffen. Ziel ist es insbesondere, dass die theoretischen Varianzen V_S^1 und V_{NR}^1 durch die Annahmen und Vereinfachungen so dargestellt werden können, dass sich möglichst viele Parameter und Terme aus dem Bruch (4.44) rauskürzen lassen. Dies gilt insbesondere für die Variationen in V_S^1 und V_{NR}^1 und wird gleich anhand der Beispiele für ausgewählte Imputationsmethoden verdeutlicht.

3. *Schätzung des Korrekturfaktors*: Es muss ein Schätzer $\hat{\alpha}_I$ für α_I gefunden werden. Hierzu müssen einige Parameter geschätzt werden. Diese hängen von der Imputationsmethode ab. Insbesondere muss die Antwortwahrscheinlichkeit p_R geschätzt werden. Unter der Annahme des uniformen Antwortmechanismus gilt hier:

$$\hat{p}_R = \frac{r}{n}. \quad (4.45)$$

Wie noch zu sehen ist, taucht in den (geschätzten) Korrekturfaktoren der Gesamtauswahlsatz f auf. Dieser wird benötigt, da wie in Abschnitt 3.4.3 erläutert, der Anteil der Varianzkomponente durch fehlende Werte insbesondere von dieser Größe abhängig ist.

4. *Verknüpfung des Korrekturfaktors mit der Resampling-Methode*: Der geschätzte Korrekturfaktor wird multiplikativ mit der Resampling-Methode verknüpft. Die Varianz wird mit Hilfe des Resampling-Verfahrens geschätzt und anschließend mit dem Korrekturfaktor multipliziert. Da $\hat{\alpha}_I$ eine Konstante darstellt, kann die Größe auch über ihre Wurzel mit den Gewichten multipliziert werden. Dies ist vor allem bei Ansätzen von Bedeutung, welche auf Replikationsgewichte abstellen.

Dieses allgemeine Vorgehen kann an Beispielen konkreter Imputationsverfahren demonstriert werden.

4.7.3.2 Mittelwertimputation

Im Rahmen der Mittelwertimputation können die Schritte zur Herleitung des Korrekturfaktors von MASHREGHI et al. (2014, S.150) wie folgt dargestellt werden (vgl. im Folgenden hierzu auch MASHREGHI et al., 2014, S.150):

1. *Varianzermittlung*: Nach MASHREGHI et al. (2014, S. 150) gilt für die Varianz der einzelnen Komponenten im Falle von Mittelwertimputation und dem Mittelwert als interessierende Statistik bei einfacher Zufallsstichprobe und der Annahme des uniformen Ausfallmechanismus und Entstehung fehlender Werte in der Grundgesamtheit:

$$V_S^1 \approx \frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2 \quad (4.46)$$

und

$$V_{NR}^1 \approx \frac{1-p_R}{p_R} \cdot \frac{1}{N^2} \sum_{i=1}^N (y_i - \bar{y}_U)^2. \quad (4.47)$$

2. *Herleitung des Korrekturfaktors*: Aus diesen beiden Komponenten kann die Konstante α_{MW} berechnet werden gemäß (4.44). Es gilt:

$$\begin{aligned} \alpha_{MW} &= \frac{V_S^1 + V_{NR}^1}{V_S^1} \quad (4.48) \\ &= \frac{\frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2 + \frac{1-p_R}{p_R} \cdot \frac{1}{N^2} \sum_{i=1}^N (y_i - \bar{y}_U)^2}{\frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2}. \end{aligned}$$

Für diesen Bruch können viele mathematische Operationen durchgeführt werden. Insbesondere lässt sich die Variation des Untersuchungsmerkmals in der Grundgesamtheit $\sum_{i=1}^N (y_i - \bar{y}_U)^2$ rauskürzen. Dies ist dem Umstand zu verdanken, dass von der einfachen Zufallsstichprobe ausgegangen wird. Bei komplexeren Stichprobendesigns kann die Stichprobenkomponente aus (4.46) nicht mehr so stark vereinfacht dargestellt werden.

Nach Durchführung der verschiedenen mathematischen Operationen, welche im Detail in Abschnitt A.1.1 im Anhang A angeben werden, resultiert aus (4.48):

$$\alpha_{MW} = (1 - p_R \cdot f) / (1 - f). \quad (4.49)$$

Bei Vorliegen des Totalwertes als interessierende Statistik resultiert der gleiche Faktor. Für diesen Schätzer werden die Varianzen der einzelnen Komponente in (4.46) und (4.47) mit N^2 multipliziert. Diese Größe kürzt sich jedoch bei der Berechnung des Korrekturfaktors im Rahmen der Quotientenbildung wieder heraus. Folglich sind die Faktoren für beide Statistiken identisch.

An Formel (4.49) lässt sich die Interpretation dieses Korrekturfaktors α_{MW} noch einmal verdeutlichen. Er beschreibt damit die Größe, um welche die gesamte Varianz die Stichprobenkomponente V_S^1 aufgrund der Berücksichtigung von V_{NR}^1 übersteigt. Wenn der Auswahlsatz f sehr klein ist, damit auch der Anteil der Varianzkomponente durch fehlende Werte, liegt α_{MW} sehr nahe bei Eins und die gesamte Varianz wird durch die Varianzkomponente V_S^1 abgedeckt. Mit immer höher werdendem Auswahlsatz wird auch (4.49) immer größer, da die Varianzkomponente bezüglich fehlender Werte V_{NR}^1 immer mehr an Gewicht gewinnt.

3. *Schätzung des Korrekturfaktors*: Um einen Schätzer (4.49) zu finden, müssen hier lediglich die Antwortwahrscheinlichkeiten p_R geschätzt werden. Der Gesamtauswahlsatz f ist im Normalfall bekannt. Damit gilt:

$$\hat{\alpha}_{MW} = (1 - \hat{p}_R \cdot f)/(1 - f). \quad (4.50)$$

Die Antwortwahrscheinlichkeit p_R wird gemäß (4.45) geschätzt.

4. *Verknüpfung des Korrekturfaktors mit der Resampling-Methode*: Die bisher vorgestellten Resampling-Methoden schätzen nur die Stichprobenkomponente der Varianz V_S^1 . Werden die berechneten Varianzschätzer jedoch mit $\hat{\alpha}_{MW}$ multipliziert, resultiert ein Varianzschätzer für die gesamte Varianz. Wie bereits beim allgemeinen Vorgehen erwähnt, kann die Multiplikation dieses Faktor entweder direkt auf die mittels des Resampling-Verfahrens ermittelte Varianz vorgenommen werden oder die Konstante $\hat{\alpha}_{MW}$ wird über ihre Wurzel in den Punktschätzer einkalkuliert. So müsste zum Beispiel beim Rescaling-Bootstrap nach CHIPPERFIELD und PRESTON (2007) aus Abschnitt 4.5 die adjustierten Gewichte multipliziert werden mit:

$$w_i^\alpha = w_i^* \cdot \sqrt{\hat{\alpha}_{MW}}. \quad (4.51)$$

Es ergibt sich der Gewichtevektor w^α und es gilt im Falle der Mittelwertimputation und der Anwendung von Reimputation und der Modifikation 2 für den Punktschätzer der Substichprobe: $\hat{\theta}_{b, resc2, reimp}^* = T(y_{M2, reimp}, w^\alpha)$.

Neben der Mittelwertimputation geben MASHREGHI et al. (2014) diese Konstanten auch für andere deterministische Imputationsverfahren an. Zum Beispiel beträgt die Konstante für Regressionsimputation bei einer verwendeten Hilfsvariable:

$$\hat{\alpha}_{Reg} = \frac{1 - \hat{p}_R \cdot f - (1 - \hat{p}_R) \cdot \widehat{\text{Cor}}_R}{(1 - f) \cdot (1 - (1 - \hat{p}_R) \cdot \widehat{\text{Cor}}_R)}. \quad (4.52)$$

$\widehat{\text{Cor}}_R$ ist der auf Basis der beobachteten Einheiten der Stichprobe berechnete Korrelationskoeffizient zwischen y und x .

4.7.3.3 Hot-Deck-Random-Imputation im Falle variabler fehlender Werte in der Grundgesamtheit

Im Falle zufälliger Imputationsverfahren sind einige Besonderheiten zu beachten. In MASHREGHI et al. (2014, S. 151) wird ein Korrekturfaktor für die (gewichtete) Hot-Deck-Random-Imputation hergeleitet. Auch dieses Vorgehen kann wieder in vier Schritte zusammengefasst werden (vgl. im Folgenden hierzu auch MASHREGHI et al., 2014, S. 151).

1. *Varianzermittlung*: Es gilt insbesondere zu beachten, dass bei zufälligen Imputationsverfahren zusätzlich noch die Imputationskomponente \tilde{V}_{Imp}^1 der Gesamtvarianz hinzukommt. Unter den gegebenen Annahmen entsprechen nach MASHREGHI et al. (2014) die ersten und zweiten Varianzkomponenten \tilde{V}_S^1 bzw. \tilde{V}_{NR}^1 denjenigen für die Mittelwertimputation aus (4.46) und (4.47). Damit gilt auch für diese Komponenten:

$$\tilde{V}_S^1 \approx \frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2 \quad (4.53)$$

und

$$\tilde{V}_{NR}^1 \approx \frac{1-p_R}{p_R} \cdot \frac{1}{N^2} \sum_{i=1}^N (y_i - \bar{y}_U)^2. \quad (4.54)$$

Für die Imputationskomponente gilt nach MASHREGHI et al. (2014):

$$\tilde{V}_{Imp}^1 \approx \frac{1-p_R}{n \cdot N} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2. \quad (4.55)$$

2. *Herleitung des Korrekturfaktors:* Auch in diesem Schritt sind die Besonderheiten zufälliger Imputationsverfahren zu berücksichtigen. Zwar werden wie bei den deterministischen Imputationsverfahren die Stichprobenkomponente \tilde{V}_S^1 grundsätzlich erfasst, die Komponente \tilde{V}_{NR}^1 hingegen nicht. Durch das Hinzukommen von \tilde{V}_{Imp}^1 ergibt sich jedoch eine Problematik. Wie bereits erörtert decken die Resampling-Verfahren bei einem kleinen Auswahlsatz f diese Komponente ab. Bei einem großen Auswahlsatz liegt jedoch eine schlechte Approximationsgüte der gemeinsamen Verteilung aus Stichproben und Donoren vor. Dies gilt für Resampling-Verfahren, welche Reskalierungen zur Berücksichtigung der Endlichkeitskorrektur verwenden. Dies gilt es ebenfalls zu korrigieren.

Nach MASHREGHI et al. (2014) ist es möglich, diesen Anteil an der Imputationskomponente zu quantifizieren, welcher durch Resampling-Verfahren mit Reskalierungen abgedeckt wird. Er beträgt für die in Abschnitt 4.7.1 in diesem Zusammenhang aufgeführten Resampling-Verfahren $(1-f) \cdot \tilde{V}_{Imp}^1$. Damit wird, wie erläutert, bei einem kleinen Auswahlsatz f die Imputationskomponente zum größten Teil abgedeckt. Bei einem sehr kleinen Auswahlsatz wird jedoch nur ein Bruchteil davon erfasst.

Insgesamt erfasst wird durch die Resampling-Verfahren der folgende Varianzterm:

$$\begin{aligned} \tilde{V}_S^1 + (1-f) \cdot \tilde{V}_{Imp}^1 &\approx \frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2 \\ &+ (1-f) \cdot \frac{1-p_R}{n \cdot N} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2. \end{aligned} \quad (4.56)$$

$\alpha_{HotDeck}$ wird wie im deterministischen Fall über den Quotienten aus der Gesamtvarianz dividiert durch den von klassischen Resampling-Verfahren abgedeckte Anteil an der Varianz berechnet. Damit wird $\alpha_{HotDeck}$ berechnet durch:

$$\alpha_{HotDeck} = \frac{\tilde{V}_S^1 + \tilde{V}_{NR}^1 + \tilde{V}_{Imp}^1}{\tilde{V}_S^1 + (1-f) \cdot \tilde{V}_{Imp}^1}. \quad (4.57)$$

Der Zähler ist damit der Summand aus den Varianzkomponenten (4.53), (4.54) und (4.55). Der Nenner ist in (4.56) gegeben. Nach einigen mathematischen Operationen, welche in Abschnitt A.1.2 im Anhang A angegeben sind, resultiert für $\alpha_{HotDeck}$:

$$\alpha_{HotDeck} = \frac{1 + p_R \cdot (1 - p_R) - p_R \cdot f}{[1 + p_R \cdot (1 - p_R)] \cdot (1 - f)}. \quad (4.58)$$

Auch hier wird wieder der Umstand ausgenutzt, dass die Variation des Untersuchungsmerkmals in der Grundgesamtheit in allen drei Varianzkomponenten vorkommt. Dadurch kann sie wieder rausgekürzt werden.

3. *Schätzung des Korrekturfaktors*: Wie im deterministischen Fall muss wieder ein Schätzer für (4.58) gefunden werden. Hierzu gilt es, die Teilnahmewahrscheinlichkeit p_R durch (4.45) zu schätzen. Der Schätzer $\hat{\alpha}_{HotDeck}$ wird anschließend berechnet durch:

$$\hat{\alpha}_{HotDeck} = \frac{(1 + \hat{p}_R \cdot (1 - \hat{p}_R)) - \hat{p}_R \cdot f}{(1 + \hat{p}_R \cdot (1 - \hat{p}_R)) \cdot (1 - f)}. \quad (4.59)$$

4. *Verknüpfung des Korrekturfaktors mit der Resampling-Methode*: Wie auch im deterministischen Fall wird die Varianz des Resampling-Verfahrens mit dem berechneten Korrekturfaktor $\hat{\alpha}_{HotDeck}$ multiplikativ verknüpft.

4.7.3.4 Hot-Deck-Random-Imputation im Falle eines fixen Response-Vektors in der Grundgesamtheit

Eine wichtige Annahme im Rahmen des Artikels von MASHREGHI et al. (2014) ist, dass der Response-Vektor in der Grundgesamtheit immer wieder von Neuem generiert wird. Im Rahmen dieser Arbeit ist jedoch auch der Fall eines fixen Response-Vektors in der Grundgesamtheit relevant. Daher wird in diesem Abschnitt eine Modifikation des beschriebenen Korrekturfaktors vorgenommen werden. Bei einem fixen Response-Vektor können die zuvor dargestellten Korrekturfaktoren nicht mehr verwendet werden. Durch die multiplikative Verknüpfung der durch Resampling-Verfahren geschätzten Varianz mit dem Korrekturfaktor wird die Varianzkomponente bezüglich fehlender Werte V_{NR}^1 bzw. \tilde{V}_{NR}^1 einkalkuliert, egal ob sie tatsächlich vorhanden ist oder nicht. Im Falle der Entstehung variabler fehlender Werte ist sie vorhanden, daher wird sie durch Berücksichtigung der Korrekturfaktoren in die Varianz auch korrekt integriert. Im Falle eines fixen Response-Vektors in der Grundgesamtheit entsteht die Komponente bezüglich fehlender Werte V_{NR}^1 bzw. \tilde{V}_{NR}^1 jedoch nicht. Hier darf der Korrekturfaktor nicht verwendet werden, da sonst die Varianz durch die nicht auftretende Varianzkomponente überschätzt wird. Es ergeben sich in Abhängigkeit der Art von Imputationsmethoden zwei verschiedene Vorgehensweisen. Im Falle deterministischer Imputationsverfahren und eines fixen Response-Vektors in der Grundgesamtheit werden Varianzen ohne Berücksichtigung von Korrekturfaktoren geschätzt. Bei zufälligen Imputationsverfahren muss wiederum eine Besonderheit Beachtung finden. Hier gilt es, die in Abschnitt 4.7.1 beschriebene schlechte Approximationsgüte der gemeinsamen Verteilung aus Stichproben und Donoren der Resampling-Verfahren bei einem großen Auswahlatz f zu korrigieren. Jedoch darf die Varianzkomponente bezüglich fehlender Werte bei einem fixen Response-Vektor nicht erfasst werden. Daher wird im

Rahmen dieser Arbeit für diesen Fall ein modifizierter Korrekturfaktor hergeleitet, welcher diese Problematik berücksichtigt. Es wird aber wieder auf die bekannten Schritte zurückgegriffen:

1. *Varianzermittlung*: Gegenüber dem Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit ist die erfassende Varianz um die Komponente \tilde{V}_{NR}^1 reduziert. Die gesamte Varianz besteht damit nur noch aus der Stichproben- und Imputationskomponente aus (4.53) und (4.55):

$$\tilde{V}_S^1 \approx \frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2 \quad (4.60)$$

und

$$\tilde{V}_{Imp}^1 \approx \frac{1-p_R}{n \cdot N} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2. \quad (4.61)$$

2. *Herleitung des Korrekturfaktors*: Wie im Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit wird für die in Abschnitt 4.7.1 in diesem Zusammenhang aufgeführten Resampling-Verfahren der Term $\tilde{V}_S^1 + (1-f) \cdot \tilde{V}_{Imp}^1$ erfasst. Es gilt:

$$\begin{aligned} \tilde{V}_S^1 + (1-f) \cdot \tilde{V}_{Imp}^1 &\approx \frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2 \\ &+ (1-f) \cdot \frac{1-p_R}{n \cdot N} \cdot \sum_{i=1}^N (y_i - \bar{y}_U)^2. \end{aligned} \quad (4.62)$$

$\alpha_{HotDeck}$ errechnet sich damit im Fall eines fixen Response-Vektors und zufälliger Imputationsverfahren gemäß:

$$\hat{\alpha}_{HotDeck} = \frac{\tilde{V}_S^1 + \tilde{V}_{Imp}^1}{\tilde{V}_S^1 + (1-f) \cdot \tilde{V}_{Imp}^1}. \quad (4.63)$$

Unter Verwendung von (4.60) und (4.61) im Zähler und (4.62) im Nenner ergibt sich nach dem üblichen Vorgehen (vgl. die Herleitung dieses Korrekturfaktors in Abschnitt A.1.3 in Anhang A) folgender Korrekturfaktor:

$$\alpha_{HotDeck} = \frac{1 + \frac{(1-p_R) \cdot p_R}{(1-f)}}{1 + (1-p_R) \cdot p_R}. \quad (4.64)$$

3. *Schätzung des Korrekturfaktors*: Die Konstante aus (4.64) wird geschätzt durch:

$$\hat{\alpha}_{HotDeck} = \frac{1 + \frac{(1-\hat{p}_R) \cdot \hat{p}_R}{(1-f)}}{1 + (1-\hat{p}_R) \cdot \hat{p}_R}. \quad (4.65)$$

4. *Verknüpfung des Korrekturfaktors mit der Resampling-Methode:* Durch multiplikative Verknüpfung der durch die mittels Resampling-Methode geschätzte Varianz und dem modifizierten Korrekturfaktor aus (4.65) kann auch im Falle eines fixen Response-Vektors in der Grundgesamtheit und zufälliger Imputationsverfahren die Varianz auch bei einem großen Auswahlstich f unverzerrt geschätzt werden. Die beschriebenen Verzerrungen werden hierdurch vermieden.

4.7.3.5 Kritische Bewertung

Im Vergleich zu der additiven Berücksichtigung der Komponente V_{NR}^1 besitzt die Verwendung von Korrekturfaktoren den Vorteil, dass ein direkter Varianzschätzer nicht separat berechnet werden muss. Daher wird die multiplikative Verknüpfung im Folgenden vorgezogen. Gegenüber der Generierung eines neuen Response-Vektors in jeder Substichprobe bietet die Verwendung von Korrekturfaktoren den Vorteil, dass das Verfahren computertechnisch weniger aufwendig ist und gleiche Korrekturfaktoren für verschiedene Resampling-Methoden verwendet werden können. Letzterer Punkt gilt zumindest für deterministische Imputationsmethoden. Beide Möglichkeiten zur Erfassung der Varianz bezüglich fehlender Werte V_{NR}^1 werden in der Simulationsstudie miteinbezogen. Die Generierung eines neuen Response-Vektors findet über den unabhängigen Rescaling-Bootstrap Berücksichtigung.

Jedoch gilt es, auch für die Verwendung von Korrekturfaktoren die Nachteile bzw. Beschränkungen zu beachten. Neben den schon beschriebenen Annahmen des uniformen Antwortmechanismus wurden diese Korrekturfaktoren für die einfache Zufallsstichprobe hergeleitet. Für komplexere Stichprobendesigns bzw. Imputationsmethoden sind diese Faktoren schwer oder gar nicht ableitbar. Das beste Beispiel sind mehrstufige Stichprobendesigns. Um z.B. $\hat{\alpha}_{MW}$ herzuleiten, müssen in (4.48) die Variationen $\sum_{i=1}^N (y_i - \bar{y}_U)^2$ in jedem Term auftauchen und vor die Klammer gezogen werden. Bei klassischen zweistufigen Stichprobendesigns beziehen sich diese Variationen auf der ersten Stufe jedoch auf die Totalwerte der PSU und auf der zweiten Stufe auf die Einheiten innerhalb der PSU. Auch müssen diese Variationen zwischen der Stichprobenkomponente und der Komponente bezüglich fehlender Werte und der Imputationskomponente nicht übereinstimmen. Die Frage, die sich stellt, ist, ob die Korrekturfaktoren, wie in diesem Abschnitt beschrieben und unter den Bedingungen der einfachen Zufallsstichprobe hergeleitet, auch für komplexere Stichprobendesigns verwendet werden können, um die Varianzkomponente bezüglich fehlender Werte bzw. die schlechte Approximationsgüte bei zufälligen Imputationsverfahren bei großen Auswahlstichen zu berücksichtigen. Dies wird in der Simulationsstudie noch näher untersucht werden. Es kann jedoch schon gesagt werden, dass ein wichtiger Faktor sein kann, wie ähnlich die Varianz des komplexen Stichprobendesigns derjenigen bei Verwendung der einfachen Zufallsstichprobe ist. Sind sich diese beiden sehr ähnlich, können die hergeleiteten Korrekturfaktoren gut zur Berücksichtigung bzw. vollständigen Erfassung der beiden Varianzkomponenten verwendet werden. In diesem Fall würde die mittels der Resampling-Methode geschätzte Varianz unter dem komplexen Design behandelt werden, als wäre sie unter der einfachen Zufallsstichprobe geschätzt worden. Wie zuvor würde sie mit dem Korrekturfaktor multipliziert werden. Liegen starke Abweichungen des komplexen Stichprobendesigns zur einfachen Zufallsstichprobe vor, kann es möglich sein, dass die Varianzkomponenten nicht gut erfasst werden.

Nach COCHRAN (1977, S. 99) können z.B. im Falle der geschichteten Zufallsstichprobe bei proportionaler Allokation große Abweichungen bei einer großen Varianz zwischen den Schichten bestehen. Dies trifft insbesondere bei sehr homogenen Schichten zu. Bei einem großen Gesamtstichprobenumfang n findet jedoch eine Annäherung statt. Hier gewinnt

auch die Varianzkomponente bezüglich fehlende Werte an Bedeutung. Homogene PSU können auch im Falle der klassischen zweistufigen Zufallsstichprobe zur Abweichung von der einfachen Zufallsstichprobe führen. Dies zeigte schon KISH (1965, S. 162) im Zusammenhang mit dem von ihm hergeleiteten Designeffekt im Fall des Ziehens mit Zurücklegen und einer gleichen PSU-Größe. Ein hoher *Intraclass-Correlation-Coefficient*, als Maßstab für homogene PSU (vgl. LOHR, 1999, S. 139), führt hier zu Unterschieden in den Varianzen der einfachen Zufallsstichprobe und der zweistufigen Zufallsstichprobe (vgl. KISH, 1965, S. 162, GANNINGER, 2010, S. 36). Des Weiteren ist auch der Einfluss einer unterschiedlichen PSU-Größe zu beachten. In diesem Zusammenhang stellt sich die Frage bei welcher Parameterkonstellation durch die Abweichung zur einfachen Zufallsstichprobe Verzerrungen entstehen. Und dieser Sachverhalt lässt sich sehr gut mit Hilfe einer Monte-Carlo-Simulation analysieren. Diese wird im nächsten Kapitel durchgeführt. Im Rahmen mehrstufiger Stichprobendesigns werden neben heterogenen auch homogene PSU, sowie gleiche und eine ungleiche PSU-Größe betrachtet, so dass die Verwendung von Korrekturfaktoren, abgeleitet auf Basis der einfachen Zufallsstichprobe, überprüft werden kann.

5 Simulation

5.1 Simulationsbeschreibung

Ziel der folgenden Simulationsstudie ist eine empirische Überprüfung der im Rahmen dieser Arbeit getroffenen theoretischen Aussagen. Die zuvor beschriebenen Schätzer bezüglich der Varianz unter fehlenden Werten und Imputation werden hinsichtlich ihrer Anwendbarkeit im Rahmen komplexer Stichprobendesigns auf ihre Eignung hin verglichen. Es erfolgt zudem eine Zerlegung der Varianz für einfachere Stichprobendesigns.

Um diese Untersuchungen durchzuführen, wird auf eine Monte-Carlo-Simulationsstudie zurückgegriffen. Aus den verschiedenen Grundgesamtheiten dieser Arbeit werden wiederholt Stichproben gezogen und die Varianzschätzer für jede dieser Stichproben berechnet. Auf diesem Wege sollen Informationen hinsichtlich ihrer Verteilung gewonnen werden. Bevor die Ergebnisse dieser Simulationsstudie präsentiert werden, wird eine ausführliche Beschreibung dieser Studie vorgenommen. Dazu werden zunächst die Grundgesamtheiten dargestellt. Diese unterscheiden sich bezüglich des verwendeten Stichprobendesigns und dem relevanten Untersuchungsgegenstand. Hinsichtlich der verwendeten Stichprobendesigns wird zwischen einer geschichteten Zufallsstichprobe und einem zweistufigen Stichprobendesign mit einer Stratifizierung auf der ersten Stufe unterschieden. Hier stehen die zuvor vorgestellten Methoden im Vordergrund. Im Falle der Varianzzerlegung werden auf stratifizierte Designs verzichtet und es werden die einfache Zufallsstichprobe und die klassische zweistufige Zufallsstichprobe verwendet. Den Abschnitt abschließen werden Informationen über die Erzeugung fehlender Werte, die Schätzer, die Szenarien und des Benchmark. Nicht alle Schätzer und Szenarien werden im Rahmen dieses Abschnittes beschrieben. Detailliertere Beschreibungen finden sich im Anhang A in den Abschnitten A.2.1 und A.2.2. Die Simulationsstudie wird mit Hilfe der statistischen Programmiersprache **R** (vgl. R CORE TEAM, 2015) umgesetzt. Als weitere Beispiele von Simulationsstudien in der Literatur in Zusammenhang mit der Varianzschätzung unter Imputation sind diejenigen in BRICK et al. (2005), MASHREGHI et al. (2014), SAIGO et al. (2001), SHAO et al. (1998), DAVISON und SARDY (2007) sowie die entsprechenden Simulationsstudien im Rahmen des DACSEIS-Projektes¹ zu nennen.

5.1.1 Einstufiges stratifiziertes Stichprobendesign

Für das einstufige stratifizierte Stichprobendesign wird der synthetische Datensatz AMELIA verwendet, welcher auf den Daten von EU-SILC 2005 basiert (vgl. im Folgenden für die Datensatzbeschreibung inklusive der einzelnen Variablen ALFONS et al., 2011 sowie KOLB, 2012). Der Datensatz besteht aus vier Regionen mit insgesamt 10.012.600 Personen bzw. 3.781.289 Haushalten. Da insbesondere die Resampling-Verfahren computertechnisch sehr aufwendig sein können, gerade unter der Berücksichtigung von Imputationen und besonders bei Untersuchungen mit großen Auswahlsätzen, findet eine Konzentration auf die Region 3 statt. Diese Region besteht aus insgesamt 821.472 Haushalten.

¹ <https://www.uni-trier.de/index.php?id=29730>

Zur Sicherstellung einer validen Varianzschätzung muss der Datensatz jedoch noch entsprechend aufbereitet werden. Die Untersuchungen werden grundsätzlich auf Haushaltsebene durchgeführt, d.h. personenbezogene Variablen müssen über die Haushaltsnummer auf Haushaltsebene transformiert werden. Dies gilt z.B. für das persönliche Einkommen. Diese Variable wird über die Mitglieder des Haushaltes zum Haushaltseinkommen aufsummiert und stellt die Untersuchungsvariable y dar. Innerhalb dieser Größe werden verschiedene persönliche Einkommenskomponenten berücksichtigt. Für die Berechnung des gesamten Haushaltseinkommens wird somit der folgende Befehl verwendet (zur Beschreibung der einzelnen Variablen, welche die Komponenten des berechneten Haushaltseinkommens bilden, vgl. ALFONS et al. (2011, S. 14) bzw. KOLB (2012, S. 182 ff.)) :

```
yk<-PY020D+PY050D+PY070D+PY090D+PY100D+PY110D+PY120D+PY130D+
  PY140D+PY010C
y<-tapply (yk ,HID ,sum)
```

Die Verteilung des Untersuchungsmerkmals y zeigt Abbildung 5.1. Die schwarz gestrichelte Linie ist das Ergebnis der Kerndichteschätzung. Es ist zu sehen, dass die Verteilung sehr rechtsschief ist, wie es bei Einkommensvariablen üblich ist. Dies stellt eine Herausforderung für die Varianzschätzung dar, gerade unter fehlenden Werten und deren Kompensation über Imputation.

Histogramm von y

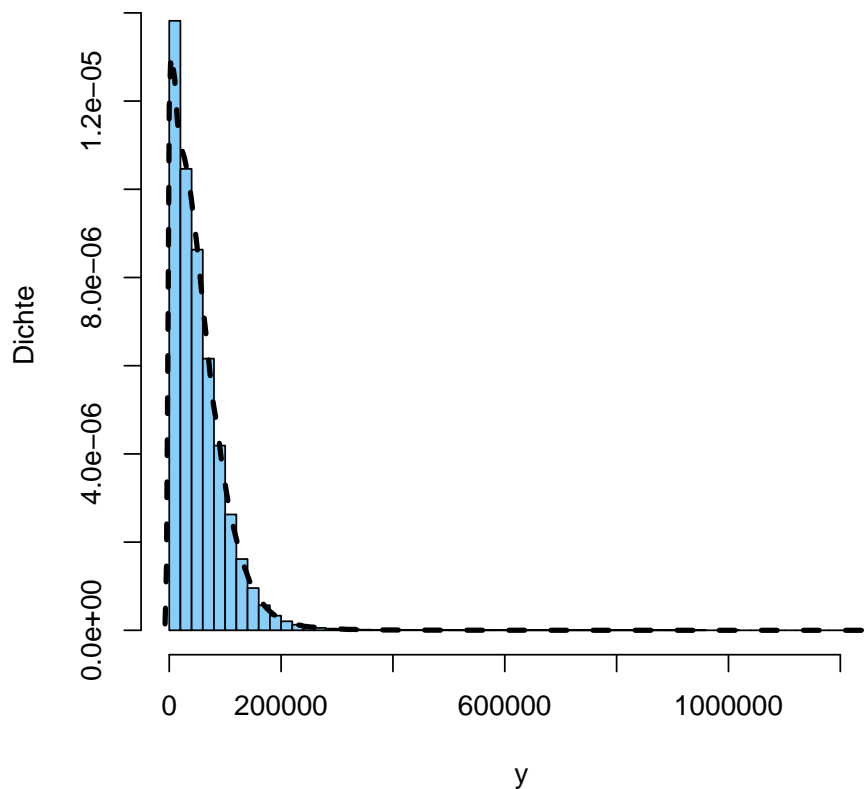


Abbildung 5.1: Histogramm und Kerndichteschätzung des Untersuchungsmerkmals im einstufigen stratifizierten Stichprobendesign

Schicht	1	2	3	4	5	6	7
Anzahl	40.489	39.811	20.211	25.128	31.477	33.382	34.848
Schicht	8	9	10	11	12	13	14
Anzahl	31.248	24.951	31.281	37.807	37.572	29.576	33.604
Schicht	15	16	17	18	19	20	21
Anzahl	28.303	37.130	29.063	35.395	26.487	32.350	31.950
Schicht	22	23	24	25			
Anzahl	31.378	35.521	36.367	46.143			

Tabelle 5.1: Schichtgrößen

Des Weiteren wird zudem eine Schichtung verwendet. Hierzu werden 25 Schichten gebildet durch Zusammenlegung von verschiedenen Gemeinden. So wird sichergestellt, dass innerhalb einer Schicht genügend Einheiten für die Varianzschätzung zur Verfügung stehen. Die ersten 24 Schichten werden aus jeweils 15 Gemeinden gebildet. Die verbleibenden Gemeinden werden Schicht 25 zugeordnet. Tabelle 5.1 gibt eine Übersicht über die daraus resultierenden Schichtgrößen in der Grundgesamtheit.

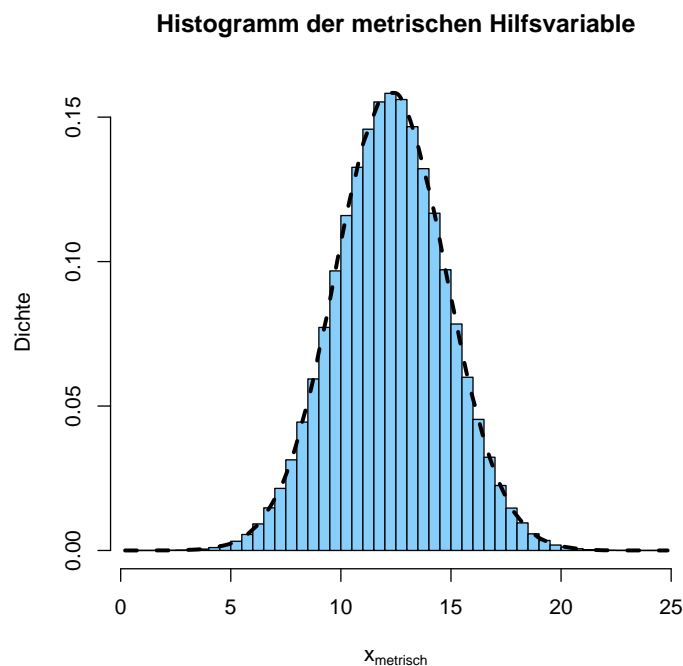


Abbildung 5.2: Histogramm und Kerndichteschätzung der metrischen Hilfsvariable im einstufigen stratifizierten Stichprobendesign

Die verwendeten Hilfsvariablen aus dem Datensatz hängen von der Imputationsmethode ab. So können beispielsweise bei der Verhältnisimputation nur metrische Variablen verwendet werden. Da im Datensatz keine metrischen Variablen vorhanden sind, wird diese über ein Modell erzeugt. Zur Erzeugung der metrischen Variablen wird das folgende Modell verwendet:

$$x_{metrisch} = (\alpha + \beta \cdot y + u) / \delta \quad (5.1)$$

wobei $u \sim N(\mu, \sigma_u)$.

Der Parameter δ wird zur Normierung der Werte auf ein bestimmtes Intervall verwendet. Dies ist notwendig, da für die Erzeugung der fehlenden Werte nach dem in Abschnitt 5.1.4 beschriebenen Vorgehen die Hilfsvariablen nicht eine zu große Bandbreite besitzen dürfen. Die Verteilung der metrischen Hilfsvariable wird in Abbildung 5.2 gegeben. Durch den großen Einfluss des Störterms ist sie normalverteilt. Der Korrelationskoeffizient der metrischen Hilfsvariable zum Untersuchungsmerkmal beträgt 0.1.

Für anderen Verfahren, wie z.B. die Hot-Deck-Random-Imputation müssen Imputationsklassen gebildet werden. Hier eignen sich insbesondere kategoriale Variablen (vgl. z.B. SHAO und SITTER, 1996, S.1281). Im Rahmen dieser Untersuchung werden zwei verschiedene kategoriale Variablen aus dem Datensatz mit einem unterschiedlich starken Zusammenhang zum Einkommen berücksichtigt: Zum einen der Haushaltstyp mit einem geringen Zusammenhang² und zum anderen die Haushaltsgröße mit einem großen Zusammenhang.³ Der Haushaltstyp besteht aus sechs Kategorien und wird in den Szenarien hauptsächlich als Hilfsvariable zur Bildung von Imputationsklassen oder im Rahmen von Predictive-Mean-Matching im Regressionsmodell verwendet. Auch die Haushaltsgröße wird auf sechs Kategorien zugrunde gelegt, indem die ursprüngliche sechste Kategorie und die darauf folgenden zu einer Kategorie zusammengefasst werden. Es soll dadurch verhindert werden, dass durch eine zu niedrige Besetzung einer Zelle bzw. Zellkombination zu wenige Elemente für eine valide Varianzschätzung zur Verfügung stehen. Diese Variable wird im Rahmen eines Szenarios bei Nearest-Neighbour-Imputations in Abwandlung zu den übrigen Szenarien als Untersuchungsmerkmal verwendet (vgl. hierzu im Detail Abschnitt 5.2.2.2). Eine genaue Übersicht über die Besetzung der einzelnen Kategorien beider Variablen wird in Tabelle 5.2 gegeben.

Kategorie	1	2	3	4	5	6
Haushaltstyp	209.990	90.672	114.442	136.762	145.594	124.012
Haushaltsgröße	169.622	233.416	178.743	169.228	50.759	19.704

Tabelle 5.2: Größe der einzelnen Kategorien

5.1.2 Mehrstufiges stratifiziertes Stichprobendesign

Im Rahmen mehrstufiger Stichprobendesigns wird eine andere Grundgesamtheit verwendet. Die Implementation dieser Grundgesamtheit in **R** orientiert sich an der Erzeugung der synthetischen Grundgesamtheit für das mehrstufige Stichprobendesign in BRUCH et al. (2011) Kapitel 4. Im Rahmen dieser Arbeit werden die für die Simulation benötigten Variablen jedoch zusammen aus einer multivariaten Normalverteilung gezogen.⁴ Dies hat insbesondere zwei Gründe.

Zum einen wirken im Rahmen von mehrstufigen Stichprobendesigns eine Mehrzahl von Einflüssen auf die Varianzschätzung. Insbesondere die Untersuchungsvariable y besitzt einen starken Einfluss. So können schiefe Verteilungen negative Einflüsse auf die Varianzschätzung besitzen. Durch diese Vielzahl von Einflüssen ist es sehr schwierig oder gar unmöglich, einzelne Einflüsse zu separieren und die Auswirkung der Variation eines bestimmten Parameters auf die Varianzschätzung zu untersuchen. Dies erschwert

² Das Eta-Quadrat liegt bei 0.0066. Dies deutet auf einen geringen Zusammenhang hin. Da der Haushaltstyp kategorial ist, muss auf dieses Maß zurückgegriffen werden.

³ Das Eta-Quadrat liegt bei 0.268

⁴ In BRUCH et al. (2011) wird das Untersuchungsmerkmal aus einer Log-Normalverteilung gezogen.

zusätzlich die Überprüfung der Wirksamkeit der unterschiedlichen Ausgestaltung einzelner Methoden auf die Varianzschätzung. Gerade dies ist jedoch das Ziel dieses Bereiches der Simulationsstudie. Verschiedene Varianten von Resampling-Methoden unter Imputation werden unter den Erfordernissen von mehrstufigen Stichprobendesigns verglichen. Durch die Wahl der Normalverteilung kann überprüft werden, ob die Methoden grundsätzlich funktionieren. Im Laufe der Simulationsstudie werden aber auch hier bestimmte Parameter bezüglich komplexerer Anforderungen variiert.

Zum anderen spielt hier insbesondere die Varianz auf den verschiedenen Stufen eine große Rolle. Im Rahmen dieser Arbeit sind auch mehrstufige Stichprobendesigns relevant, bei denen nicht die gesamte Varianz auf der ersten Stufe liegt. In manchen Szenarien ist es wichtig, dass ein möglichst großer Anteil der Varianz auf den nachfolgenden Stufen vorhanden ist. Solche Szenarien lassen sich gut über die Erstellung der Grundgesamtheit aus einer Normalverteilung bewerkstelligen, durch geeignete Wahl ihrer Parameter. Zur Implementation eines mehrstufigen Stichprobendesigns ist es weiterhin notwendig, dass eine ausreichend große Grundgesamtheit zur Verfügung steht. Daher wurde ihre Größe auf 2.500.000 Einheiten auf der letzten Stufe festgelegt, wobei von einem zweistufigen Design mit Stratifizierung auf der ersten Stufe ausgegangen wird.

Damit gilt für den Datensatz $\zeta = (y, x_1, x_2)$, wobei y weiterhin das Untersuchungsmerkmal und x_1 und x_2 die beiden Hilfsmerkmale sind:

$$\zeta \sim N(\mu, \Sigma), \quad (5.2)$$

wobei für die Parameter μ und Σ gilt:

$$\mu = (1.000, 500, 800) \quad (5.3)$$

und

$$\Sigma = \begin{pmatrix} 500 & 130 & 70 \\ 130 & 200 & 70 \\ 70 & 70 & 150 \end{pmatrix}. \quad (5.4)$$

Die Verteilung des Untersuchungsmerkmals ist in Abbildung 5.3 zu sehen. Die Bestimmung der Parameter liegen verschiedene Kriterien zugrunde. Es wird für die Beobachtungsvariable y eine große Varianz gewählt, so dass auch auf der zweiten Stufe genügend Varianz vorhanden ist. Daher ist diese Varianzkomponente am größten. Die Varianzen innerhalb der Hilfsvariablen sind deutlich geringer. Dies hängt mit der Erzeugung fehlender Werte in Abschnitt 5.1.4 zusammen. Hilfsvariablen mit großer Varianz erschweren die Anwendung der dort beschriebenen Modelle. Die Kovarianzen zwischen den Variablen dürfen nicht zu hoch gewählt werden, da für die mehrstufigen Designs die Analyse des Ausfallmechanismus MAR im Vordergrund steht. Bei höherer Korrelation zwischen einer bzw. mehrerer Hilfsvariablen zum Untersuchungsmerkmal, wäre auch ein Einfluss des Untersuchungsmerkmals auf die Generierung der fehlenden Werte gegeben und der Ausfallmechanismus hätte sich somit in Richtung NMAR verlagert. Der Korrelationskoeffizient der ersten Hilfsvariable zum Untersuchungsmerkmal beträgt 0.41 und derjenige der zweiten Hilfsvariable zum Untersuchungsmerkmal 0.25.

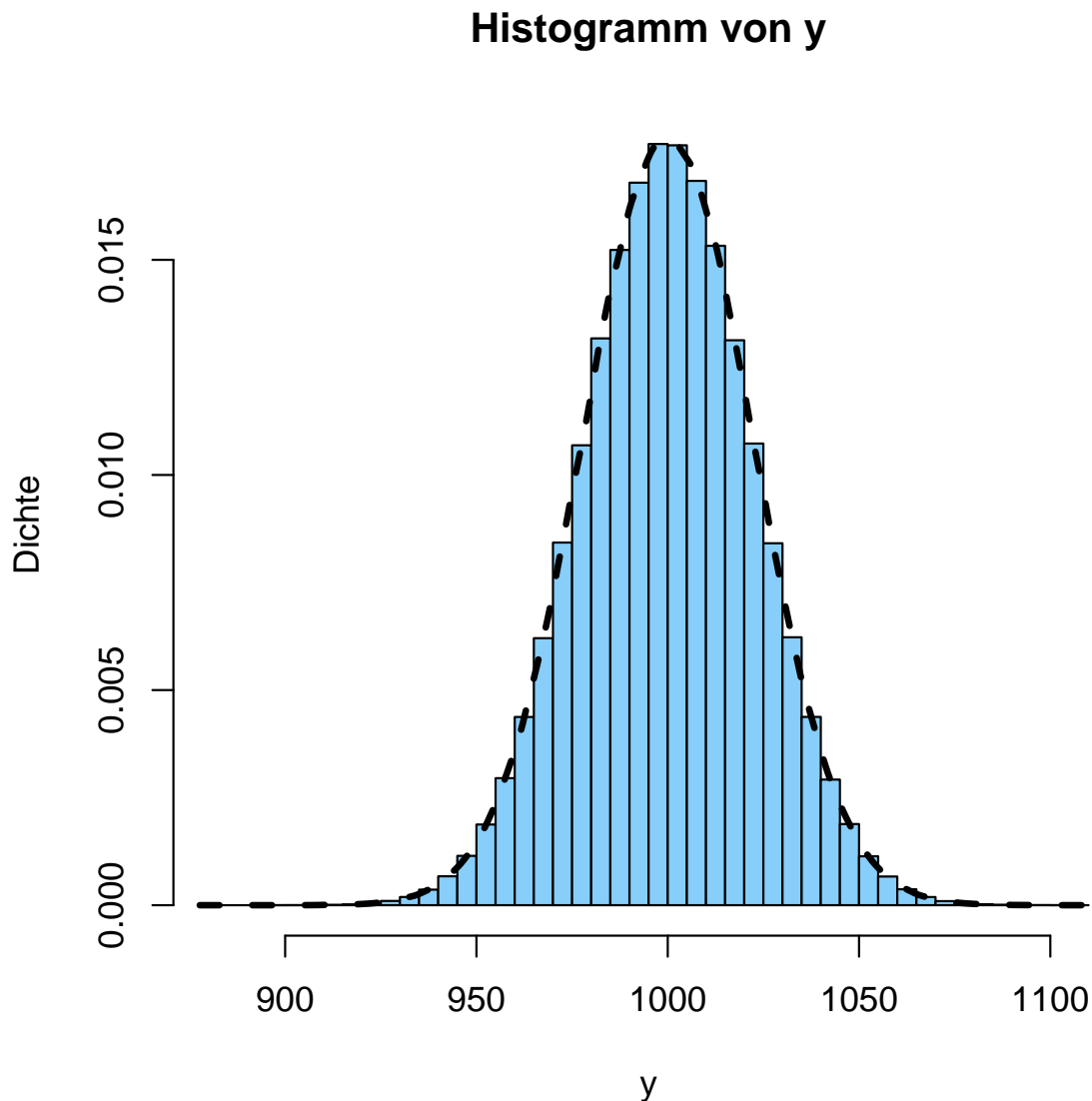


Abbildung 5.3: Histogramm und Kerndichteschätzung des Untersuchungsmerkmals im Szenario des mehrstufigen Stichprobendesigns

Im Rahmen der Simulationsstudie für das mehrstufige Stichprobendesign werden hauptsächlich einfache Imputationsmethoden wie Mittelwertimputation und Hot-Deck-Random-Imputation verwendet. Die Begründung liegt auch hier in der Vielzahl an Einflüssen des mehrstufigen Stichprobendesigns und der Gewährleistung der Untersuchung des Einflusses eines bestimmten Parameters bzw. eines originären Methodenvergleichs. Für Imputationsverfahren wie Mittelwertimputation und Hot-Deck-Random-Imputation ist es jedoch wieder notwendig, eine kategoriale Hilfsvariable zu erzeugen, um die Imputationsklassen zu bilden. Die Grundlage hierfür ist die metrische Hilfsvariable x_1 . Den Einheiten wird anhand der Ausprägungen dieser Variable die Kategorie 1 bis 6 zugeordnet, in dem Sinne, dass sich Einheiten mit einer ähnlich hohen Ausprägung bezüglich x_1 in einer Kategorie befinden. Aufgrund dieses Generierungsprozesses besitzt die kategoriale Hilfsvariable einen gewissen, aber nicht zu hohen Zusammenhang zum Untersuchungsmerkmal. Das Zusammenhangsmaß Eta-Quadrat liegt hier bei 0.155. Die Besetzung der einzelnen Kategorien ist in Tabelle 5.3 dargestellt. Diese kategoriale Variable wird für diese beiden

Kategorie	1	2	3	4	5	6
Besetzung	417.004	416.618	416.931	416.565	416.574	416.308

Tabelle 5.3: Größe der einzelnen Kategorien im mehrstufigen Design

Imputationsmethoden als alleinige Hilfsvariable in den betreffenden Szenarien benutzt. In einem großen Teil der Szenarien sind die einzelnen Schichten und die einzelnen PSU gleich groß. Dadurch liegt auch ein großer Anteil der Varianz auf der zweiten Stufe. Bei einer stark unterschiedlichen PSU-Größe liegt ein großer Anteil der Varianz auf der ersten Stufe und die gewünschten Untersuchungen können nicht mehr durchgeführt werden. Außerdem soll zunächst einmal der Einfluss einer unterschiedlichen PSU-Größe ausgeklammert werden, um einen originären Vergleich der Methoden zu gewährleisten. Daher besteht in einem großen Anteil der Szenarien jede Schicht aus 50 PSU und jede PSU aus 2.000 USU. Wie im Fall der einstufigen geschichteten Zufallsstichprobe werden 25 Schichten verwendet. Damit ergeben sich insgesamt 1250 PSU.

Jedoch ist auch die Untersuchung der Auswirkung einer unterschiedlichen PSU-Größe auf die Varianzschätzung wichtig. In bestimmten Szenarien werden daher auch eine unterschiedliche Schicht- und PSU-Größe berücksichtigt. Die Wahl der Größe der PSU und der Schichten wurde dahingehend vorgenommen, dass eine ausreichende Anzahl an PSU innerhalb der Schichten und USU innerhalb der PSU zur validen Durchführung der Varianzschätzung zur Verfügung stehen. Innerhalb der Simulationsstudie werden zwei Szenarien unterschieden, in denen die Schicht- und PSU-Größe unterschiedlich stark variiert werden. Bei kleinerer Variation liegt die Schichtgröße zwischen 46 und 52 PSU pro Schicht und zwischen 1.900 und 2.100 USU pro PSU. Im Falle einer größeren Variation resultiert eine Schichtgröße zwischen 40 und 60 PSU pro Schicht und eine PSU-Größe zwischen 1.000 und 3.000 USU pro PSU. Die Befehle in **R** zur Erzeugung der Schichten und PSU im Falle stark unterschiedlicher PSU-Größen werden in Abschnitt A.2.3 im Anhang A angegeben.

Des Weiteren wird in der Simulationsstudie noch zwischen heterogenen und homogenen PSU unterschieden. Bei heterogenen PSU unterscheiden sich die Einheiten innerhalb der PSU sehr stark. Es liegt eine große Varianz innerhalb der PSU vor und damit auch auf der zweiten Stufe. Dies ist bei Verwendung des normalverteilten Untersuchungsmerkmals der Fall. Bei homogenen PSU sind sich die Einheiten innerhalb der PSU sehr ähnlich. Hier liegt die Varianz zu einem großen Anteil zwischen den PSU und damit auf der ersten Stufe des Stichprobendesigns. Homogene PSU werden in der Simulationsstudie durch die Sortierung des normalverteilten Untersuchungsmerkmals nach der Höhe seine Ausprägung erreicht.

5.1.3 Varianzzerlegung

In Rahmen dieser Simulationsstudie wird der Betrachtung der einzelnen Komponenten der Varianz, wie sie in Abschnitt 3.4.2 behandelt worden sind, besondere Aufmerksamkeit geschenkt. Genauer gesagt werden die einzelnen Komponenten aus (3.23), (3.24), (3.25), (3.26), (3.27) und (3.28) für das ein- und mehrstufige Stichprobendesign mittels einer Monte-Carlo-Simulation berechnet und dargestellt. In einem weiteren Schritt werden für die wichtigsten Szenarien im einstufigen Stichprobendesign gezeigt, welche dieser Komponenten einzelne Modifikationen des Rescaling-Bootstrap ohne Zurücklegen genau abdecken. Diese Resampling-Methode ist von besonderer Bedeutung, da im Rahmen die-

ser Arbeit ein Ansatz entwickelt worden ist, wie dieses Verfahren auf den Fall fehlender Werte und deren Kompensation über Imputation übertragen werden kann.

Um den Beitrag einzelner Komponenten fehlerfrei darzustellen, ist es unabdingbar, andere Einflüsse, wie z.B. die Schiefe der Verteilung des Untersuchungsmerkmals oder zusätzliche Auswirkungen komplexer Imputationsmethoden und Stichprobendesigns, möglichst gering zu halten. Daher wird als Grundgesamtheit diejenige aus dem mehrstufigen Design aus Abschnitt 5.1.2 verwendet. Hier ist das Untersuchungsmerkmal normalverteilt. Bezüglich des Stichprobendesigns wird zur Komplexitätsreduktion auf eine Stratifizierung verzichtet. Es wird im einstufigen Fall die einfache Zufallsstichprobe und im mehrstufigen Fall ein zweistufiges Stichprobendesign ohne Stratifizierung auf der ersten Stufe verwendet. Des Weiteren findet eine Konzentration auf einfache Imputationsmethoden wie Mittelwert- und Hot-Deck-Random-Imputation statt. Zur Gewährleistung der Qualität dieser Methoden werden sie wie zuvor innerhalb der Imputationsklassen angewendet.

Ziel dieser Varianzzerlegung ist es, weiterhin auch den Einfluss bestimmter Parameter zu untersuchen. Hierbei geht es insbesondere um die Überprüfung der theoretischen Aussagen aus Abschnitt 3.4.3. Diese Untersuchung muss *ceteris paribus* vorgenommen werden. Es wird von Szenario zu Szenario ein Parameter, wie z.B. der Auswahlsatz, die Rate oder der Fall der Entstehung fehlender Werte sowie die Imputationsmethode variiert. Alle anderen Parameter werden konstant gehalten. Nur auf diesem Weg ist der Einfluss eines Parameters zu bestimmen.

Wichtig ist in diesem Zusammenhang, wie die Varianzzerlegung simulativ umgesetzt wird. Hier werden die inneren und äußeren Vorgänge der Varianzkomponenten aus (3.23), (3.24), (3.25), (3.26), (3.27) und (3.28) unter Verwendung von Programmierschleifen reproduziert. Die Varianz wird über die Durchführung einer Monte-Carlo-Simulation zerlegt (*Monte-Carlo-Zerlegung der Varianz*).⁵ Dieser Vorgang wird im Folgenden am Beispiel der Imputationskomponente \tilde{V}_{Imp}^1 aus (3.26), im Falle der Entstehung variabler fehlender Werte in der Grundgesamtheit und Hot-Deck-Random-Imputation erläutert. Sie ist in der Varianzzerlegung (3.26) definiert durch $E_q E_p V_I(\hat{\theta}_I | s, z)$. Diese Komponente lässt sich, wie in Abschnitt 3.4.2 beschrieben, über drei Vorgänge darstellen und simulativ wie folgt umsetzen. Der innerste Vorgang ist der Imputationsprozess. Gegeben einer konkreten Stichprobenrealisation und eines gegebenen Response-Vektors wird der Imputationsprozess 100-mal wiederholt und für den erhaltenen imputierten Vektor und mit den entsprechenden Designgewichten die interessierende Statistik geschätzt. Anschließend wird über diese 100 Schätzer die Varianz berechnet. Der direkt übergeordnete Vorgang ist die Stichprobenziehung. Auch diese wird 100-mal wiederholt. Für jede dieser 100 Wiederholungen wird der innerste Vorgang der Berechnung der bedingten Imputationsvarianz vollzogen. Es resultieren damit 100 dieser Varianzen, über die der Mittelwert berechnet wird. Der äußerste Vorgang ist die Generierung des Response-Vektors. Dieser Prozess wird 1.000-mal wiederholt und in allen diesen Wiederholungen die beiden zuvor beschriebenen Vorgänge durchgeführt. Auch über die 1.000 erhaltenen Erwartungswerte wird wieder ein Mittelwert gebildet, welcher die Monte-Carlo-Approximation Varianzkomponente \tilde{V}_{Imp}^1 darstellt. Es resultieren dadurch für eine Komponente im Fall zufälliger Imputationsverfahren $1.000 \cdot 100 \cdot 100 = 10.000.000$ Wiederholungen. Daher können für die inneren Vorgänge mit nur 100 Wiederholungen gearbeitet werden. Wie noch zu sehen sein wird, sind die dadurch gewonnenen Schätzungen der Komponenten über die hohe Anzahl der Wiederholun-

⁵ Auch in MASHREGHI et al. (2014) in Abschnitt 7 wird eine Varianzzerlegung durchgeführt, jedoch in einem deutlich geringeren Umfang und für die geschichtete Zufallsstichprobe. Dort wird auch ein anderer Ansatz für die Varianzzerlegung gewählt. MASHREGHI et al. (2014) schätzen die einzelnen Komponenten über eine Linearisierung.

gen insgesamt sehr stabil. Vereinfacht, unter Auslassung bestimmter Komplexitäten des Imputationsprozesses, lässt sich der Programmiercode für die Imputationskomponente im Fall variabler fehlender Werte in der Grundgesamtheit und Hot-Deck-Random-Imputation in **R** wie folgt darstellen:

```
EW.Imp.Var <- rep(NA,1000)

for(i in 1:1000){

##### Generierung des Response Vektors z #####

  z <- nr_gen(IKL)

##### IKL beschreibt die Variable, nach der die
##### Imputationsklassen gebildet werden.
##### Basierend auf dieser Variable wird der Ausfall erzeugt.

  Var.pest.imp <-rep(NA,100)

for(u in 1:100){

  n <- 1250000
  nr <-c(1:length(y))

##### Stichprobenziehung durch einfache Zufallsstichprobe #####

  S.ID <- sample(nr,size=n)
  S.ID <- as.vector(unlist(S.ID))

  z_s <- z[S.ID]
  y_s <- y[S.ID]

  IKL_s <- IKL[S.ID]
  gew_s <- gew[S.ID] ##### gew ist der Gewichtevektor

  p.est.imp <- rep(NA,100)

for(j in 1:100){

##### Anwendung der Imputationsmethode #####

  Imp <- tapply(c(1:length(IKL_s)), IKL_s, function(x,l,u,p,y)
    HotDeck(y_s[x],z_s[x]))

##### HotDeck ist die Funktion, welche
##### die Hot-Deck-Random-Imputation durchfuehrt.

  y_imp <- unlist(sapply(Imp,function(x){y<-x$yNA.imputed}))
  y_s[!z_s] <- y_imp
```

```
##### Berechnung des imputierten Schaetzers #####

  p.est.imp[j] <- statistic(y_s, gew_s)

##### Die Funktion statistic berechnet die
##### interessierende Statistik, hier den Totalwert.

}
##### Berechnung der Varianz der 100 imputierten Schaetzer

  Var.pest.imp[u] <- var(p.est.imp)

}
##### Berechnung des Erwartungswertes der 100 Varianzen #####

  EW.Imp.Var[i] <- mean(Var.pest.imp)

}

##### Berechnung der Imputationskomponente #####

  Bed.Imp.Var <- mean(EW.Imp.Var)
```

Auch die beiden anderen Komponenten werden gemäß dieses Vorgehens umgesetzt, wobei hier die Erwartungswerte und Varianzen in anderer Reihenfolge gebildet werden, wie Formel (3.26) zeigt. Im Falle der Entstehung fehlender Werte in der Stichprobe werden im äußersten Vorgang die Stichproben gezogen, wogegen der untergeordnete Vorgang die Erzeugung des Response-Vektors ist, wie in Abschnitt 3.4.2 theoretisch erläutert wurde und in (3.27) dargestellt ist. Bei den deterministischen Imputationsverfahren sind nur zwei Komponenten mit jeweils zwei Vorgängen relevant. Daher können hier für den inneren Vorgang statt 100 Wiederholungen 1.000 verwendet werden. Im Fall eines festen Response-Vektors entfällt die Schätzung der Komponente bezüglich fehlender Werte.

Für mehrstufige Stichprobendesigns wird wie erläutert ein zweistufiges Design ohne Schichtung verwendet. Es wird hier nur der Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit betrachtet. Fehlende Werte werden hier auf der letzten Stufe gebildet und es wird ausschließlich die Mittelwertimputation verwendet. Daher ist die Varianz auf der ersten Stufe des Stichprobendesigns vollständig der Stichprobenkomponente V_S^1 zuzurechnen. Diese wird daher berechnet, indem zunächst für einen gegebenen Response-Vektor wiederholt die im Szenario vorgesehene Anzahl von PSU gezogen wird. Es werden alle Einheiten der gezogenen PSU berücksichtigt und fehlende Werte durch die Mittelwertimputation und basierend auf Imputationsklassen imputiert. Der interessierende Schätzer wird auf diesen imputierten Daten für jede der Wiederholungen berechnet. Anschließend erfolgt die Varianzbildung über die Schätzer der einzelnen Wiederholungen. Diese Varianzberechnung wird für jeden einzelnen der generierten Response-Vektoren durchgeführt und anschließend der Erwartungswert über die verschiedenen Varianzen gebildet. Er resultiert die Stichprobenkomponente $V_{S_1}^1$ der ersten Stufe. Für beide Vorgänge werden 1.000 Wiederholungen angesetzt. Dieser Varianz der ersten Stufe wird diejenige Varianz V_1 gegenübergestellt, welche auf Basis von vollständigen Beobachtungen berechnet worden ist. Dadurch soll die Auswirkung der Imputation der Einheiten der letzten Stufe auf

die Varianz der ersten Stufe verdeutlicht werden.

Die Varianzzerlegung auf der zweiten und damit letzten Stufe ist derjenigen aus der einfachen Zufallsstichprobe im Falle der Entstehung variabler fehlender Werte in der Grundgesamtheit sehr ähnlich. Es fallen beide Varianzkomponenten V_S^1 und V_{NR}^1 an, welche auf der zweiten Stufe durch V_{S2}^1 und V_{NR2}^1 bezeichnet werden. Es gibt jedoch einen Unterschied zur einfachen Zufallsstichprobe. Der innere Vorgang wird zusätzlich gegeben der Realisationen der ersten Stufe durchgeführt. So wird z.B. für die Stichprobenkomponente V_{S2}^1 für einen gegebenen Response-Vektor und eine gegebene Realisation von PSU auf der ersten Stufe die Stichprobenziehung auf der zweiten Stufe wiederholt und entsprechend der interessierende imputierte Schätzer berechnet. Über die erhaltenen Schätzer wird die Varianz gebildet. Auf der übergeordneten Stufe werden wiederholt PSU gezogen und für jede erhaltene Stichprobenrealisation von PSU der beschriebene Vorgang der Varianzberechnung durchgeführt. Über die erhaltenen Varianzen wird der Erwartungswert gebildet. Dieser Vorgang wird für verschiedene Response-Vektoren wiederholt und über die erhaltenen Größen der Erwartungswert berechnet. Es resultiert die Komponente V_{S2}^1 . Aufgrund dieses Vorgehens wird die Zerlegung computertechnisch sehr aufwendig und nur für den Fall eines deterministischen Imputationsverfahrens durchgeführt. Hier werden für die beiden inneren Vorgänge 100 Wiederholungen und für den äußeren Vorgang 1.000 Wiederholungen veranschlagt. Dadurch resultieren schon für den deterministischen Fall 10.000.000 Wiederholungen.

5.1.4 Erzeugung der fehlenden Werte

Unabhängig von dem verwendeten Datensatz bzw. dem Stichprobendesign müssen die fehlenden Werte erzeugt werden. Auch für den im einstufigen Design verwendeten Datensatz AMELIA liegen in den interessierenden Variablen keine fehlenden Werte vor.

Zudem bietet die Erzeugung fehlender Werte den großen Vorteil der Überprüfung des Einflusses des Antwortmechanismus auf die Varianzschätzung. Es wird die Möglichkeit gegeben, verschiedene Szenarien zu bilden, welche sich hinsichtlich Parameter wie der Ausfallrate, dem Antwortmechanismus oder der Stelle der Entstehung fehlender Werte unterscheiden.

Das genaue Vorgehen der Generierung fehlender Wert hängt in diesem Zusammenhang insbesondere vom betrachteten Antwortmechanismus ab. Im Rahmen dieser Arbeit wird insbesondere die in 2.1 beschriebene Einteilung der Antwortmechanismen in MCAR, MAR und NMAR zugrunde gelegt.

Im Falle von MCAR werden in Höhe einer bestimmten Ausfallrate zufällig Einheiten (im einstufigen Design z.B. die Haushalte) entweder aus der Grundgesamtheit oder der Stichprobe gezogen, welche bezüglich des Untersuchungsmerkmals y einen fehlenden Wert besitzen. Für alle verwendeten Szenarien wird angenommen, dass für die Hilfsvariablen x_k für $k = 1, \dots, K$ keine fehlenden Werte vorliegen.

Für die Fälle von MAR und NMAR wird auf Antwortwahrscheinlichkeiten der einzelnen Haushalte abgestellt, welche nach einem logistischen Modell berechnet werden (vgl. im Folgenden ENDERLE et al., 2013, S. 95). Liegt z.B. MAR mit einer metrischen Hilfsvariable x_1 vor, gilt für die Berechnung der Antwortwahrscheinlichkeiten der Einheit i :

$$p_{Ri} = \frac{\exp(\tilde{\alpha} + \tilde{\beta} \cdot x_{1i})}{1 + \exp(\tilde{\alpha} + \tilde{\beta} \cdot x_{1i})}. \quad (5.5)$$

Die Parameter $\tilde{\alpha}$ und $\tilde{\beta}$ werden z.B. in Abhängigkeit der zu erzielenden Ausfallrate und der Stärke des Zusammenhanges des Response-Vektors z mit der metrischen Hilfsvariablen

x_1 bestimmt. Sie unterscheiden sich zwischen den Szenarien und die jeweiligen Werte werden in der Szenarienbeschreibung im Anhang A in Abschnitt A.2.2 angegeben. Ist die Hilfsvariable kategorial statt metrisch, erfolgt die Berechnung nach:

$$p_{Ri} = \frac{\exp(\tilde{\alpha} + \tilde{\gamma}_i)}{1 + \exp(\tilde{\alpha} + \tilde{\gamma}_i)}. \quad (5.6)$$

$\tilde{\gamma}_i$ beschreibt einen Parameter, dessen Wert für die Einheit i sich dahingehend bestimmt, in welcher Kategorie der kategorialen Variablen x_2 sich i befindet. Eine Kategorie wird in diesem Zusammenhang auf Null gesetzt. In manchen Szenarien werden fehlende Werte auch bezüglich zweier Variablen generiert. Handelt es sich hierbei z.B. um eine metrische und kategoriale Variable gilt:

$$p_{Ri} = \frac{\exp(\tilde{\alpha} + \tilde{\beta} \cdot x_{1i} + \tilde{\gamma}_i)}{1 + \exp(\tilde{\alpha} + \tilde{\beta} \cdot x_{1i} + \tilde{\gamma}_i)}. \quad (5.7)$$

Im Falle von MAR werden für die Erzeugung fehlender Werte in den meisten Szenarien auf Hilfsvariablen zurückgegriffen, welche zum Untersuchungsmerkmal einen nicht zu großen Zusammenhang aufweisen. Der Grund liegt darin, dass bei einem großen Zusammenhang zu den Hilfsvariablen auch das Untersuchungsmerkmal Einfluss auf die Erzeugung fehlender Werte besitzt. Dadurch kann sich der Ausfallmechanismus schnell in Richtung NMAR ändern. Ein Ausfallmechanismus MAR wäre unter Umständen nicht mehr gegeben.

Wird der Antwortmechanismus zu NMAR geändert, hängt die Antwortwahrscheinlichkeit von i nicht mehr von den Hilfsvariablen, sondern vom Untersuchungsmerkmal y ab. Da dieses Merkmal im Rahmen dieser Arbeit größtenteils metrisch ist, gilt:

$$p_{Ri} = \frac{\exp(\tilde{\alpha} + \tilde{\beta} \cdot y_i)}{1 + \exp(\tilde{\alpha} + \tilde{\beta} \cdot y_i)}. \quad (5.8)$$

Nach der Berechnung der Antwortwahrscheinlichkeiten wird im folgenden Schritt der Response-Vektor z gebildet und damit festgelegt, ob eine Einheit bezüglich y einen beobachteten oder einen fehlenden Wert besitzt. Eine Einheit i antwortet, wenn gilt $p_{Ri} > \omega_i$, wobei ω_i aus einer Gleichverteilung $U(0, 1)$ gezogen wird. Aufgrund dieses Vorgehens variiert die Ausfallrate von Response-Vektor-Erzeugung zu Response-Vektor-Erzeugung. Die Differenzen sind bei einer gegebenen Parameterkonstellation gering. Jedoch kann im Folgenden die Ausfallrate nicht exakt angegeben werden.

Die genaue Verteilung und der Zusammenhang des Response-Vektors mit den wichtigsten Variablen für die einstufige geschichtete Zufallsstichprobe und das mehrstufige Stichprobendesign mit einer Stratifizierung auf der ersten Stufe wird in den Abbildungen 5.4 und 5.5 wiedergegeben. In beiden Abbildungen sind jeweils die Fälle der Erzeugung fehlender Werte in der Grundgesamtheit zum einen bezüglich der erzeugten metrischen Variable und zum anderen hinsichtlich der kategorialen Variable enthalten. Eine detaillierte Beschreibung der Variablen wurde in den Abschnitten 5.1.1 und 5.1.2 gegeben. Die Ausprägung TRUE bzw. 1 des Response-Vektors steht für eine antwortende Person, FALSE oder NULL für eine Einheit mit fehlendem Wert. In Abbildung 5.4 befindet sich die Verteilung des Response-Vektors im Zusammenhang mit der metrischen Hilfsvariable im linken oberen Quadranten. Die rote Linie kennzeichnet die Unterscheidung der Mittelwerte der Kategorien Beobachtet und Fehlend. Es ist zu sehen, dass fehlende Werte eher mit einem höheren

Wert der Hilfsvariable auftreten. Es handelt sich damit um einen Ausfallmechanismus MAR. Der gleiche Sachverhalt gilt für die Erzeugung fehlender Werte nach der metrischen Hilfsvariable im mehrstufigen Design mit einer Stratifizierung auf der ersten Stufe, dargestellt in der linken Hälfte in Abbildung 5.5. Werden fehlende Werte bezüglich der kategorialen Hilfsvariable erzeugt (im linken unteren Quadranten in Abbildung 5.4 bzw. in der rechten Hälfte von Abbildung 5.5), ist zu sehen, dass der Ausfall deutlich größer in den Kategorien 4, 5 und 6 ist, repräsentiert durch den roten Balken. Des Weiteren werden im einstufigen Design noch die Verteilung des Response-Vektors unter MCAR und NMAR dargestellt. Bei ersterem ist der Ausfall in allen Kategorien gleich verteilt. Bei NMAR ist der Ausfall größer mit steigendem Einkommen, welches das Untersuchungsmerkmal präsentiert.

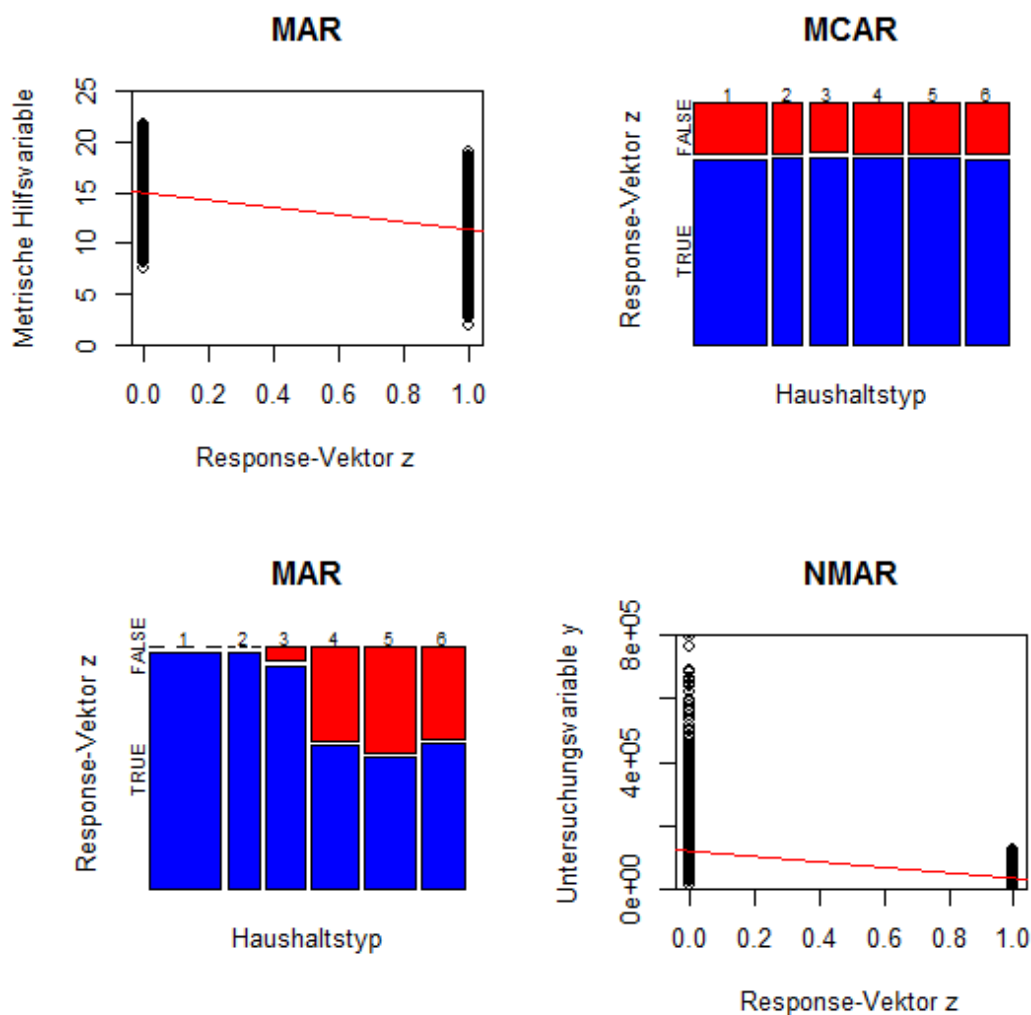


Abbildung 5.4: Verteilung fehlender Werte für die beiden Hilfsvariablen im einstufigen stratifizierten Design

Entscheidend ist zudem, wo genau dieser Prozess der Generierung fehlender Werte durchgeführt wird. Wie zuvor beschrieben, können fehlende Werte ein einziges Mal oder wiederholt in der Grundgesamtheit sowie nach der Stichprobenziehung gebildet werden. Dies hat einen entscheidenden Einfluss auf die Varianzschätzung. Daher sollen im Rahmen der Simulationsstudie alle drei dieser Möglichkeiten berücksichtigt werden.

Im Falle einer einmaligen Erzeugung fehlender Werte in der Grundgesamtheit wird die zuvor beschriebene Berechnung der Antwortwahrscheinlichkeiten und die Auswahl der antwortenden bzw. nicht antwortenden Einheiten auf alle Einheiten in der Grundgesamtheit durchgeführt. Es resultiert ein einzelner Response-Vektor z , welcher in jedem Simulationsdurchlauf zugrunde gelegt wird. Anschließend werden in Abhängigkeit des Stichprobendesigns die jeweiligen Einheiten zur Berechnung der interessierenden Statistik in jedem Simulationsdurchlauf gezogen. Im Rahmen der Entstehung variabler fehlender Werte in der Grundgesamtheit wird hingegen der Prozess der Erzeugung des Response-Vektors z in jedem Simulationsdurchgang vor der Stichprobenziehung auf Basis der Einheiten der Grundgesamtheit durchgeführt. In jedem Simulationsdurchgang wird daher ein neuer Response-Vektor z in Länge der Grundgesamtheit erzeugt. Im Fall der Entstehung fehlender Werte in der Stichprobe findet zunächst in jedem Simulationsdurchlauf die Stichprobenziehung statt. Anschließend werden für die gezogenen Einheiten nach den beschriebenen Modellen die Antwortwahrscheinlichkeiten berechnet und die antwortenden Einheiten nach dem beschriebenen Verfahren ermittelt. Es resultiert auch hier in jedem Simulationsdurchlauf ein neuer Response-Vektor z , aber in Länge der Stichprobengröße. Dieser Vorgang wird unabhängig bestimmter Besonderheiten der Stichprobenrealisationen durchgeführt und hängt nur von der Hilfsvariablen ab. Die erhaltenen Response-Vektoren sind daher dem Fall variabler fehlender Werte in der Grundgesamtheit sehr ähnlich.

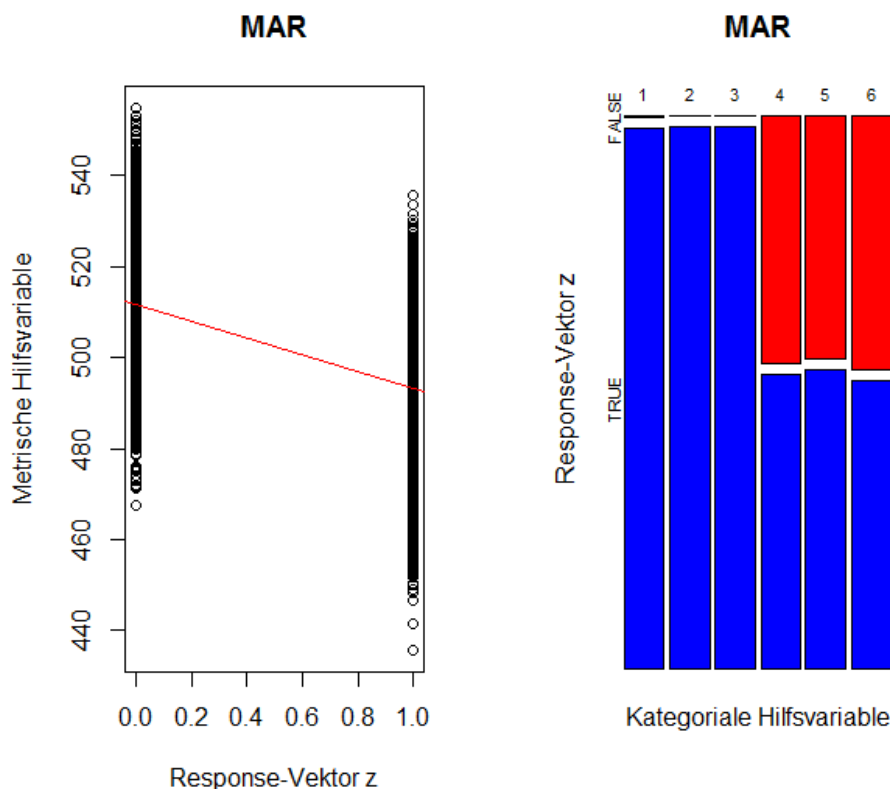


Abbildung 5.5: Verteilung fehlender Werte für die beiden Hilfsvariablen im mehrstufigen Design mit einer Stratifizierung auf der ersten Stufe

5.1.5 Szenarien

Die einzelnen Szenarien unterscheiden sich sehr stark bezüglich der Ausprägung der einzelnen Parameter. Daher kann an dieser Stelle nur ein kurzer Überblick über die wichtigsten Parameter gegeben werden. Eine ausführliche Darstellung erfolgt im Anhang A in Abschnitt A.2.2. Beispiele solcher Parameter, welche variiert werden, sind die Imputationsmethode, der Fall der Entstehung fehlender Werte, der Ausfallmechanismus, die Stichprobengröße (bzw. der Auswahlatz) oder die Ausfallrate. Die ersten drei Parameter und ihre Ausprägungen wurden im Rahmen der Arbeit ausführlich erläutert. Für den Auswahlatz und die Ausfallrate gelten folgende allgemeine Erläuterungen.

Die einzelnen Szenarien unterscheiden sich bezüglich der Stichprobengröße. Im einstufigen stratifizierten Stichprobendesign werden zwei Auswahlätze verwendet.⁶ Im Rahmen kleiner Auswahlätze werden insgesamt 40.000 Einheiten gezogen. Es ergibt sich ein Gesamtauswahlatz von fast 5 %. Zudem wird ein größerer Gesamtauswahlatz von ungefähr 46 % mit einer Stichprobengröße von 380.000 Einheiten verwendet. Die Stichprobenumfänge werden hier proportional auf die Schichten verteilt. Für das mehrstufige Design mit Schichtung auf der ersten Stufe erfolgt auf der ersten Stufe eine Gleichaufteilung des Stichprobenumfangs und auf der zweiten Stufe eine proportionale Aufteilung auf die PSU.⁷ Es werden drei verschiedene Stichprobenumfänge verwendet. Im Rahmen der kleinsten Stichprobengröße werden in jeder Schicht $l_h = 5$ PSU bei einem Gesamtstichprobenumfang von 25.000 Einheiten gezogen. Bei einem mittleren Stichprobenumfang werden $l_h = 10$ Einheiten auf der ersten Stufe gezogen und der Gesamtstichprobenumfang beträgt 125.000 Einheiten. Im Rahmen eines großen Auswahlatzes werden $l_h = 38$ PSU in jeder Schicht ausgewählt und der Gesamtstichprobenumfang liegt bei 625.000 Einheiten. Für die Varianzzerlegung im Zusammenhang mit der einfachen Zufallsstichprobe werden drei Stichprobenumfänge verwendet: 40.000 (kleiner Auswahlatz von 1,6 %), 125.000 (mittlerer Auswahlatz von 50 %) und 2.250.000 (hoher Auswahlatz von 90 %). Für die Varianzzerlegung im mehrstufigen Design wird auf folgende Umfänge abgestellt: $l = 125$ PSU auf der ersten Stufe und ein Gesamtstichprobenumfang von 25.000 Einheiten auf der zweiten Stufe sowie $l = 950$ gezogene PSU und ein Gesamtstichprobenumfang von 1.250.000 Einheiten.

Die Ausfallrate beträgt insgesamt im einstufigen stratifizierten Stichprobendesign zwischen 21 % und 24 %. Für Szenarien, in denen der Response-Vektor über die metrische Hilfsvariable erzeugt wird, liegt sie etwas höher als bei Generierung der fehlenden Werte über die kategoriale Variable. Bei der Erzeugung nach dem kategorialen Hilfsmerkmal im Falle von MAR unterscheidet sich die Ausfallrate zwischen den sechs Kategorien deutlich, mit Werten zwischen 0 % und ca. 45 %. Diese Ausfallraten werden in einzelnen Kategorien und auch insgesamt in manchen Szenarien zur Überprüfung der Auswirkungen eines höheren Ausfalls auf die Varianzschätzung deutlich erhöht. Im mehrstufigen Stichprobendesign mit einer Stratifizierung auf der ersten Stufe liegt sie bei Erzeugung fehlender Werte nach der kategorialen Hilfsvariable bei ungefähr 22 %, bei Erzeugung nach der metrischen Hilfsvariable bei ca. 36 %. Bei letzterem sollen die Auswirkungen

⁶ Die genaue betragsmäßige Definition von kleinen, mittleren und großen Auswahlätzen bzw. Stichprobengrößen ist vom Szenario und dem beabsichtigten Untersuchungsgegenstand abhängig. Die Szenarien sind jedoch so konstruiert, dass im Falle eines größeren Gesamtauswahlatzes f auch die Auswahlätze innerhalb jeder Schicht f_h größer sind. Daher werden Szenarien durch die allgemeine Bezeichnung kleinere, mittlere und große Auswahlätze abgegrenzt.

⁷ Auch hier werden die Szenarien so konstruiert, dass mit einem größeren Gesamtauswahlatz f sowohl größere Auswahlätze an PSU innerhalb der Schichten f_h wie auch größere Auswahlätze innerhalb der PSU f_{hd} einhergehen.

höherer Ausfallraten überprüft werden. Die Erzeugung fehlender Werte nach der metrischen Hilfsvariable spielt im Rahmen der Simulationsstudie im mehrstufigen Fall nur eine untergeordnete Rolle. Auch im mehrstufigen Design können, bei Verwendung der kategorialen Variablen, die Ausfallraten innerhalb der Kategorien in der Größenordnung des einstufigen Designs schwanken. Im Falle der Varianzzerlegung für die einfachen Zufallsstichprobe liegt sie bei einer kleinen Ausfallrate bei ungefähr 5 %, bei der mittleren Ausfallrate bei ungefähr 22,5 % und bei ungefähr 43 % im Rahmen der großen Ausfallrate. Für das zweistufige Design liegt die Rate bei ungefähr 22,5 %.

5.1.6 Schätzer

Die interessierende Statistik stellt der Totalwert dar. Als Punktschätzer wird der Horvitz-Thompson-Schätzer verwendet, wie in Formel (3.1) bzw. Abschnitt 3.1 definiert.

Im Rahmen der Simulation werden verschiedene, der in den vorangegangenen Kapitel theoretisch erörterte Varianzschätzer berücksichtigt. Welche der Methoden genau in den einzelnen Szenarien zur Anwendung kommen, hängt insbesondere von der zugrunde liegenden Imputationsmethode ab, aber auch die Laufzeiten der einzelnen Verfahren sind entscheidend. Gerade Resampling-Verfahren gehen mit längeren Rechenzeiten einher. Eine Berücksichtigung aller in den vorherigen Kapiteln und Abschnitten beschriebenen Varianzschätzmethode in allen Szenarien ist daher nicht sinnvoll.

Problematisch stellt sich im Rahmen der Varianzschätzer die Einhaltung der Annahme des uniformen Antwortmechanismus dar. Für verschiedene Methoden (vgl. Abschnitt 3.6 und 4.1.2) sollte dieser zumindest innerhalb der Imputationsklassen gewährleistet sein. Dies ist bei bestimmten Imputationsmethoden in Abhängigkeit des Ausfallmechanismus nicht möglich. Werden fehlende Werte in Abhängigkeit einer metrischen Variable erzeugt, welche eine Vielzahl von Werten zulässt, insbesondere auch nicht ganzzahlige Realisationen, können die Antwortwahrscheinlichkeiten sehr unterschiedlich sein. Dies ist in der Simulationsstudie in den MAR-Szenarien der Fall, bei denen die nach dem Modell (5.1) erzeugte Variable dem Ausfallmechanismus zugrunde liegt. Dies betrifft vor allem die Szenarien mit Verhältnisimputation, deterministischer wie stochastischer Regressionsimputation, Nearest-Neighbour-Imputation und Predictive-Mean-Matching. Die Variable bzw. die Variablen nach denen fehlende Werte erzeugt worden sind, werden grundsätzlich auch in das Imputationsmodell integriert. Eine ausschließliche Abstimmung auf kategoriale Variablen zur Gewährleistung des uniformen Ausfallmechanismus für die fünf aufgezählten Imputationsmethoden wäre aus unterschiedlichen Gründen nicht sinnvoll. Verhältnisimputation setzt metrische Variablen voraus. Bei ausschließlicher Verwendung von kategorialen Variablen, im Rahmen der deterministischen und stochastischen Regressionsimputation, wäre die Variation der unabhängigen Variable zu gering. Dies gilt insbesondere, wenn die verwendete kategoriale Variable auch zur Bildung der Imputationsklasse verwendet und Regressionsimputation innerhalb der Klasse durchgeführt wird. Hier würde gar keine Variation entstehen. Dieser Sachverhalt gilt auch für Predictive-Mean-Matching. Diese Methode basiert ebenfalls auf einem Regressionsmodell. Nearest-Neighbour-Imputation ist bei Verwendung von kategorialen Variablen der Hot-Deck-Random-Imputation basierend auf Imputationsklassen sehr ähnlich. Der Grund liegt darin, dass viele Einheiten mit gleicher Distanz vorliegen können, aus welchen Zufallsstichproben gezogen werden müssen. Die daraus entstehenden Problematiken werden schon im Rahmen von Hot-Deck-Random-Imputation behandelt. Daher wird auch bei Nearest-Neighbour-Imputation, die nach dem Modell (5.1) erzeugte metrische Variable, zur Distanzermittlung benutzt. Durch die Verwendung dieser metrischen Variable zur Erzeugung fehlender Werte in diesen Szenarien

kann die Annahme des uniformen Ausfallmechanismus nicht aufrechterhalten werden. Wie noch zu sehen sein wird, sind hierdurch aber insbesondere die direkten Varianzschätzer und weniger die Resampling-Verfahren betroffen.

Aus diesen Erläuterungen ergibt sich ebenfalls die Bildung der Imputationsklassen. Diese werden, wenn ihre Verwendung erfolgt, in den meisten Szenarien über die Schichten hinweg gebildet. Die Ausnahmen werden nachfolgend erläutert. Da die Annahme des uniformen Ausfallmechanismus nicht aufrecht erhalten werden kann, wird bei der deterministischen und stochastischen Regressionsimputation sowie bei der Verhältnisimputation auf die Bildung von Imputationsklassen verzichtet. Die Hilfsvariablen gehen hier über das Imputationsmodell ein. In Abschnitt 2.2 wurden bereits erwähnt, dass für Mittelwertimputation und Hot-Deck-Random-Imputation in ihrer ursprünglichen Version keine Hilfsinformationen verwendet werden. Zur Begegnung dieses Nachteils werden diese Verfahren innerhalb von Imputationsklassen angewendet.⁸ Wie in den Abschnitten 5.1.1 und 5.1.2 beschrieben, werden diese über die kategoriale Variable Haushaltstyp gebildet. Damit liegen sechs Klassen vor. Auch für die Fractional-Imputation, welche in dieser Arbeit als eine Form von der Hot-Deck-Random-Imputation benutzt wird, werden aus diesem Grund Imputationsklassen und damit die kategoriale Hilfsvariable verwendet. Gleiches gilt für Nearest-Neighbour-Imputation. Wie hier noch zu sehen sein wird, resultieren für dieses Verfahren sehr variable Varianzschätzer. Um diese Variabilität etwas einzudämmen, werden Imputationsklassen durch eine synthetisch erzeugte Variable gebildet, welche zu der metrischen Hilfsvariable aus (5.1) einen starken Zusammenhang besitzt.⁹ Sie dient als Distanzmaß. Außerdem sind Nearest-Neighbour-Verfahren bei großen Stichprobenumfängen sehr computerintensiv aufgrund der Berechnung der Distanzmatrix. Durch die Bildung von Imputationsklassen und der Berechnung von Distanzmatrizen innerhalb der Klassen wird eine große Distanzmatrix vermieden. Aus diesem Grund wird auch für Predictive-Mean-Matching auf Imputationsklassen abgestellt, wobei hierzu die Schichtungsvariable genommen wird. Gegenüber Nearest-Neighbour-Imputation wird das kategoriale Hilfsmerkmal im zugrunde liegenden Regressionsmodell verwendet.

Imputationen innerhalb der Berechnung des Punktschätzers werden daher gemäß dieser Vorgehensweise und in Abhängigkeit der Imputationsmethode in der Imputationsklasse oder auf der gesamten Stichprobe durchgeführt. Für die Varianzschätzer wird dies bestmöglich implementiert. Kritisch ist in diesem Zusammenhang, dass für die geschichteten Stichprobendesigns die Imputationsklassen und Schichten nicht übereinstimmen. Die direkten Varianzschätzer aus Abschnitt 3.6 werden innerhalb der Schichten durchgeführt. Dadurch kann es im Falle von MAR zu Annahmeverletzungen dieser Verfahren kommen. Wie in Abschnitt 3.6 erläutert, setzen die direkten Varianzschätzer häufig MCAR oder den uniformen Antwortmechanismus voraus. Eine gleichzeitige Berücksichtigung der Schichten und der Imputationsklassen ist für viele der direkten Schätzer sehr schwierig oder gar nicht möglich. Für Resampling-Verfahren kann die Nichtübereinstimmung von Schichten und Imputationsklasse in der Durchführung berücksichtigt werden. Dies wurde schon in Kapitel 4 beschrieben. Grundsätzlich werden die Substichproben innerhalb der Schichten und die Reimputationen bzw. Adjustierungen im Falle des Vorhandenseins von

⁸ Im Rahmen der Simulationsstudie wird für Hot-Deck-Random-Imputation ausschließlich die ungewichtete Ziehung von Spendern verwendet. Durch die Berücksichtigung unterschiedlicher Auswahlwahrscheinlichkeiten wird das Verfahren sehr computerintensiv. Die Spender werden außerdem mit Zurücklegen innerhalb der Imputationsklassen gezogen.

⁹ Diese Hilfsvariable zur Bildung der Imputationsklassen bei Nearest-Neighbour-Imputation wird nicht über ein Modell erzeugt, sondern die Einheiten werden entsprechend ihrer Ausprägung der metrischen Hilfsvariablen den einzelnen Kategorien zugeordnet. Aufgrund der beschriebenen Variabilität werden hier 78 Imputationsklassen gebildet. Das Eta-Quadrat zur metrischen Hilfsvariable beträgt 0.99.

Imputationsklassen in diesen Kategorien durchgeführt. Eine Ausnahme bildet der Delete-a-Group-Jackknife. Hier haben Voruntersuchungen gezeigt, dass bessere Resultate erzielt werden können, bei Anwendung der Reimputationen bzw. Adjustierungen innerhalb der Schichten. Ein Grund hierfür ist seine schichtbezogene Anwendung wie in (4.17) definiert. Eine andere Ausnahme bildet der Rescaling-Bootstrap aus Abschnitt 4.7.3. MASHREGHI et al. (2014) gehen bei dessen Herleitung von übereinstimmenden Schichten und Imputationsklassen mit gleichen Antwortwahrscheinlichkeiten innerhalb der Kategorien aus. Hier muss geschaut werden, ob diese Annahmeverletzung bezüglich der Anwendung negative Auswirkungen hat.

Des Weiteren basieren die in Abschnitt 4.7.3 dargestellten Korrekturfaktoren zur Berücksichtigung der Varianzkomponente bezüglich fehlender Werte sowie der schlechten Approximationsgüte bei zufälligen Imputationsverfahren und großen Gesamtauswahlsätzen auf der einfachen Zufallsstichprobe und uniformen Antwortmechanismus. Diese müssen in der Simulationsstudie auch bei komplexen Stichprobendesigns berücksichtigt werden, um die im Rahmen dieser Arbeit wichtigen Aussagen bei großen Auswahlsätzen, insbesondere über das Auftreten bestimmter Varianzkomponenten und die Anwendung von Resampling-Verfahren, zu überprüfen. Hierzu wird untersucht, ob sich die Verwendung dieser Korrekturfaktoren auch eignet, wenn keine einfache Zufallsstichprobe und kein uniformer Antwortmechanismus vorliegt.

In Abschnitt 4.1.2 wurde die Unterscheidung zwischen Reimputation und Adjustierung der imputierten Werte in jeder Substichprobe vorgenommen. Bei deterministischen Imputationsverfahren können beide Vorgehensweisen angewendet werden, da sie zu den gleichen Ergebnissen führen. Aufgrund der einfacheren Umsetzung wird im Rahmen der Simulation in diesem Fall auf die Reimputation abgestellt und daher in den entsprechenden Abschnitten von Reimputation gesprochen. Bei zufälligen Imputationsverfahren geht die Reimputation, wie beschrieben mit Überschätzungen der Varianz einher. Für diese Verfahren wird daher die Adjustierung der imputierten Werte angewendet und in den entsprechenden Abschnitten erfolgt die Verwendung dieses Begriffes. Davon unabhängig wird im Rahmen von Hot-Deck-Random-Imputation auch ein Vergleich zwischen den beiden Ansätzen durchgeführt, um die Überschätzung der Reimputation aufzudecken.

Aus Gründen der Vergleichbarkeit sollten Resampling-Verfahren auf einer einheitlichen Anzahl an Substichproben basieren. Dies wird auch in dieser Arbeit angestrebt. Im Falle der einstufigen geschichteten Zufallsstichprobe ist dieser Sachverhalt unproblematisch. Die einzelnen Parameter der Resampling-Verfahren werden dahingehend bestimmt, dass den einzelnen Resampling-Verfahren 125 Substichproben zugrunde liegen. Bei den Bootstrap-Verfahren werden B Replikationen entsprechend gewählt. Beim BRR kann die Anzahl über die wiederholte Anwendung nach RAO und SHAO (1996) aus Abschnitt 4.6.1 erreicht werden. Beim Delete-a-Group-Jackknife werden 5 Gruppen in jeder Schicht gewählt, womit insgesamt 125 Substichproben resultieren. Bei der Random-Group-Methode werden 125 Gruppen über die Schichten hinweg gebildet. Bei den mehrstufigen Stichprobendesigns werden nur der Rescaling-Bootstrap und der BRR berücksichtigt. Auch hier gelten die entsprechenden Erläuterungen aus dem einstufigen Design. Es liegen 125 Substichproben zugrunde. Jedoch stellt die Anwendung des BRR auf der zweiten Stufe eine Ausnahme dar. Diese Methode wird gemäß dem in den Abschnitten 4.6.2 und 4.6.3 beschriebenen Vorgehen umgesetzt. Die PSU werden in diesem Fall wie Schichten behandelt. Die Anzahl der für die Stichprobe gezogenen PSU der einzelnen Szenarien bestimmt hier die Anzahl der Substichproben. Unter Berücksichtigung der wiederholten Anwendung des BRR nach RAO und SHAO (1996) ergeben sich für den BRR auf der zweiten Stufe deutlich mehr als 125 Substichproben. Dies gilt es bei den Interpretationen zu beachten. Eine einheit-

liche Substichprobenanzahl ist im Falle mehrstufiger Stichprobendesigns sehr schwierig einzuhalten.

Die für die Simulation benötigten Schätzer wurden direkt in **R** umgesetzt. In diesem Zusammenhang wurde für die deterministische und stochastische Regressionsimputation sowie den BRR auf die **R**-Codes im Anhang von DAVISON und SARDY (2004b) zurückgegriffen. Der BRR wird jedoch in dieser Arbeit unter Imputation wie in Abschnitt 4.6.3 beschrieben, verwendet. Dementsprechend mussten die Funktionen aus DAVISON und SARDY (2004b) angepasst werden. Auch die **R**-Codes für die beiden Imputationsmethoden mussten bezüglich der Gegebenheiten der Simulationsstudie abgewandelt werden.

Eine Übersicht, inklusive einer genauen Bezeichnung und kurzen Beschreibung der einzelnen Varianzschätzmethoden dieser Arbeit, wird in den Tabellen in Abschnitt A.2.1 im Anhang A gegeben.

5.1.7 Benchmark Problematik

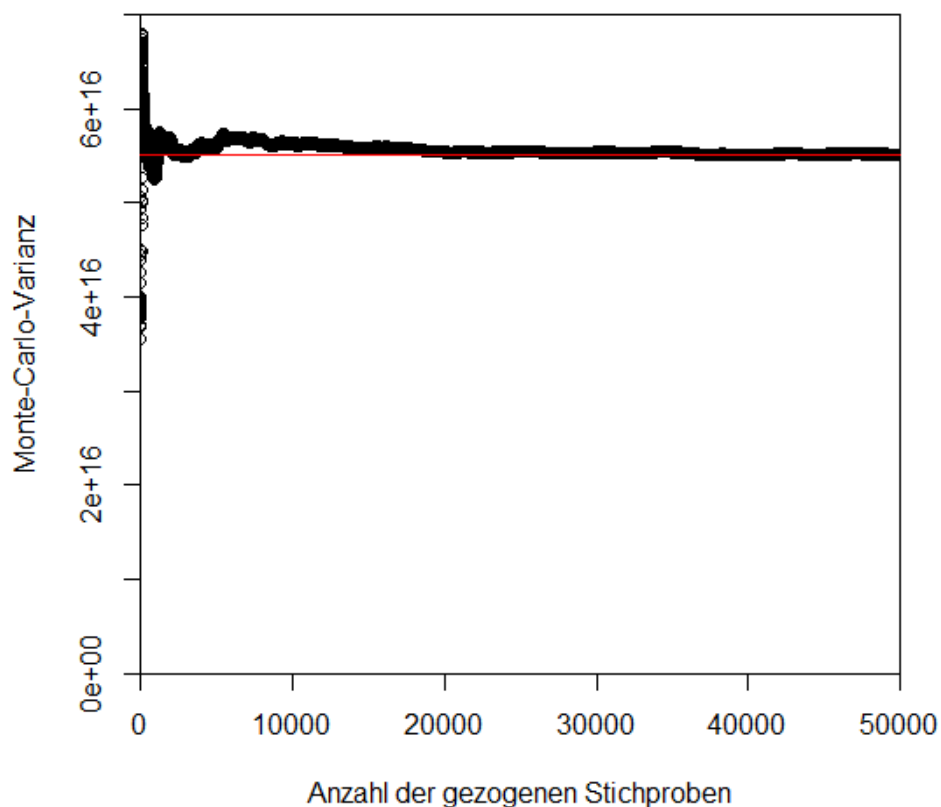


Abbildung 5.6: Stabilität der Monte-Carlo-Varianz für den Fall von Mittelwertimputation, kleinen Auswahlsätzen und der Entstehung variabler fehlender Werte in der Grundgesamtheit

Als Benchmark im Rahmen von Simulationen in Zusammenhang mit Varianzschätzungen, insbesondere wenn die wahre Varianz schwer oder nicht ermittelt werden kann, dient die Monte-Carlo-Varianz. Dies ist im Rahmen von Imputationen häufig der Fall, vor allem

wenn eine sehr komplexe Imputationsmethode vorliegt. Auch in dieser Arbeit wird auf die Monte-Carlo-Varianz als Benchmark zurückgegriffen. Diese Varianz stellt genauer gesagt die Varianz der verschiedenen Punktschätzer dar, welche im Rahmen der Monte-Carlo-Simulation für die einzelnen aus der Grundgesamtheit gezogenen Stichproben berechnet werden. Die hierzu notwendige Anzahl an Stichproben hängt sehr stark vom Anwendungsfall ab. Um die Stabilität des Benchmarks zu gewährleisten, ist es oft notwendig die Berechnung der Monte-Carlo-Varianz auf 10.000 Stichproben vorzunehmen. Dies verdeutlichen die folgenden Abbildungen 5.6 und 5.7.

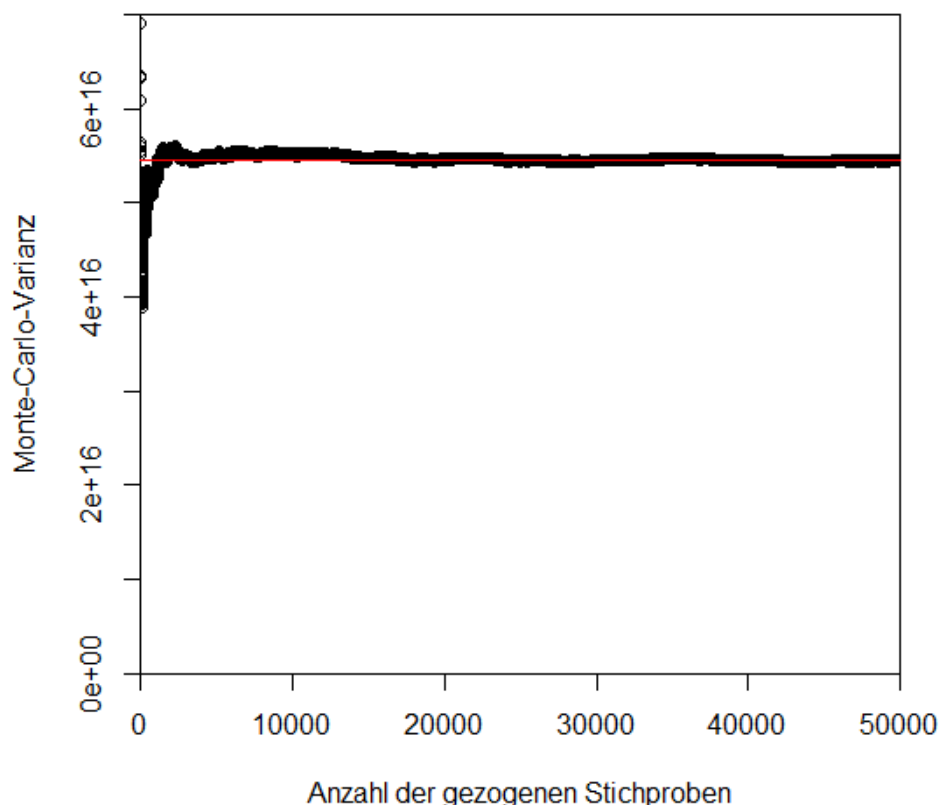


Abbildung 5.7: Stabilität der Monte-Carlo-Varianz für den Fall von Mittelwertimputation, kleinen Auswahlsätzen und der Entstehung fehlender Werte in der Stichprobe

In Abbildung 5.6 wird die Erzeugung fehlender Werte wiederholt in der Grundgesamtheit vorgenommen und in Abbildung 5.7 in der Stichprobe. Beiden Abbildungen liegt Mittelwertimputation zugrunde. Die Abbildungen zeigen die Stabilität des Schätzers in Abhängigkeit der Anzahl der gezogenen Stichproben. Es lässt sich feststellen, dass bis zu 10.000 Wiederholungen bei beiden Fällen keine Stabilität gegeben ist. Die rote Linie beschreibt in beiden Fällen den Wert der Monte-Carlo-Varianz bei 50.000 Wiederholungen. Es ist zu sehen, dass im Fall der Entstehung fehlender Werte in der Grundgesamtheit bis zu einer Anzahl von 10.000 Simulationsdurchläufe die Monte-Carlo-Varianz zum Teil sehr hoch über der roten Linie liegt. Im Falle der Entstehung in der Stichprobe nähert sich die Monte-Carlo-Varianz zunächst von unten. Eine Stabilität wird in beiden Fällen erst

zwischen 10.000 und 20.000 Wiederholungen erreicht. Jedoch kommt es auch hier noch zu kleineren Veränderungen. Durch diese Instabilitäten ist es möglich, dass die Monte-Carlo-Varianz beider Fälle für eine gleiche Anzahl an zugrunde liegenden Stichproben geringfügig voneinander abweichen, obwohl eine Übereinstimmung erwartet wird.

Als Konsequenz dieser Instabilitäten wird die Monte-Carlo-Varianz im Rahmen der Varianzzerlegung bei einstufigen -und mehrstufigen Stichprobendesigns ohne Schichtung auf Basis von 10.000 Wiederholungen berechnet. Liegen komplexere Stichprobendesigns mit Schichtungen vor, werden sogar 20.000 Stichproben verwendet. Bei einer besonders instabilen Monte-Carlo-Varianz in einigen wenigen Szenarien wird sogar auf 50.000 erhöht. Im Gegensatz zum Benchmark konvergieren die Varianzschätzmethoden meist deutlich schneller. Des Weiteren sind insbesondere die Resampling-Verfahren mit größeren Laufzeiten verbunden. Daher werden die Varianzschätzmethoden auf 1.000 Stichproben berechnet.

5.2 Ergebnisse bezüglich des einstufigen stratifizierten Stichprobendesigns

In den nachfolgenden Abschnitten werden die Ergebnisse der Simulationsstudie analysiert. Zunächst werden die Resultate bezüglich des einstufigen Stichprobendesigns dargestellt. Es folgt die Interpretation der Ergebnisse bezüglich der Varianzzerlegung im einstufigen Design. Der letzte Teil der Ergebnisinterpretation der Simulationsstudie bildet die Darstellung der Resultate für die mehrstufigen Stichprobendesigns.

Bezüglich des einstufigen stratifizierten Stichprobendesigns erfolgt zunächst die Darstellung der Ergebnisse verschiedener in den vorangegangenen Abschnitten behandelter Varianzschätzmethoden. Hier steht der Methodenvergleich im Vordergrund. Es werden einfachere Imputationsmethoden wie Mittelwertimputation zugrunde gelegt. Anschließend erfolgt die Betrachtung ausgewählter Varianzschätzmethoden für komplexere Imputationsmethoden wie Nearest-Neighbour-Imputation. Der darauf folgende Abschnitt zeigt die Ergebnisse der Anwendung von Varianzschätzern hinsichtlich komplexerer Anwendungsfälle wie die Kombination von Imputationsmethoden, höherer Ausfallraten oder der Variation des Ausfallmechanismus. Der Abschnitt der Darstellung der Simulationsergebnisse des einstufigen Stichprobendesigns schließt mit den für diese Arbeit so wichtigen Fällen größerer Auswahlätze.

Wie in Abschnitt 5.1.7 beschrieben, werden die Varianzschätzmethoden für 1.000 Stichproben berechnet. Die Verteilungen dieser Schätzergebnisse der einzelnen Varianzschätzmethoden für das einstufige Stichprobendesign werden durch Boxplots visualisiert. Der Mittelwert der 1.000 Schätzergebnisse wird durch einen schwarzen Punkt im Boxplot repräsentiert. Die rote Linie stellt die in Abschnitt 5.1.7 beschriebene Monte-Carlo-Varianz dar, welche als Benchmark verwendet wird.

5.2.1 Vergleich verschiedener Varianzschätzmethoden

Ziel dieses Abschnitts ist der Vergleich verschiedener in den vorherigen Kapiteln und Abschnitten beschriebenen Methoden bezüglich ihrer Eignung der Schätzung der Varianz bei Vorhandensein von fehlenden Werten und der Anwendung von Imputationen. Die meisten Verfahren liegen für Mittelwertimputation, Verhältnisimputation, Regressionsimputation und Hot-Deck-Randomimputation vor. Daher bezieht sich der Vergleich auf diese Imputationsmethoden.

5.2.1.1 Mittelwertimputation

Zunächst wird der Fall der Mittelwertimputation betrachtet.¹⁰ Abbildung 5.8 gibt die Ergebnisse der Varianzschätzung im Falle kleiner Auswahlsätze wieder. Fehlende Werte wurden in jedem Simulationsdurchlauf in der Grundgesamtheit erzeugt. Es wurden in diesem Szenario keine Korrekturfaktoren zur Erfassung der Varianz bezüglich fehlender Werte berücksichtigt, mit Ausnahme der unabhängigen Rescaling-Bootstrap-Methoden und des direkten Varianzschätzers, bei denen die Erfassung immanent ist. Die Varianz bezüglich fehlender Werte ist aufgrund des geringen Gesamtauswahlsatzes vernachlässigbar gering. Fast alle Methoden sind entweder unverzerrt oder überschätzen die Varianz. Daher kann hier auf die Erfassung dieser Komponente der Varianz verzichtet werden.

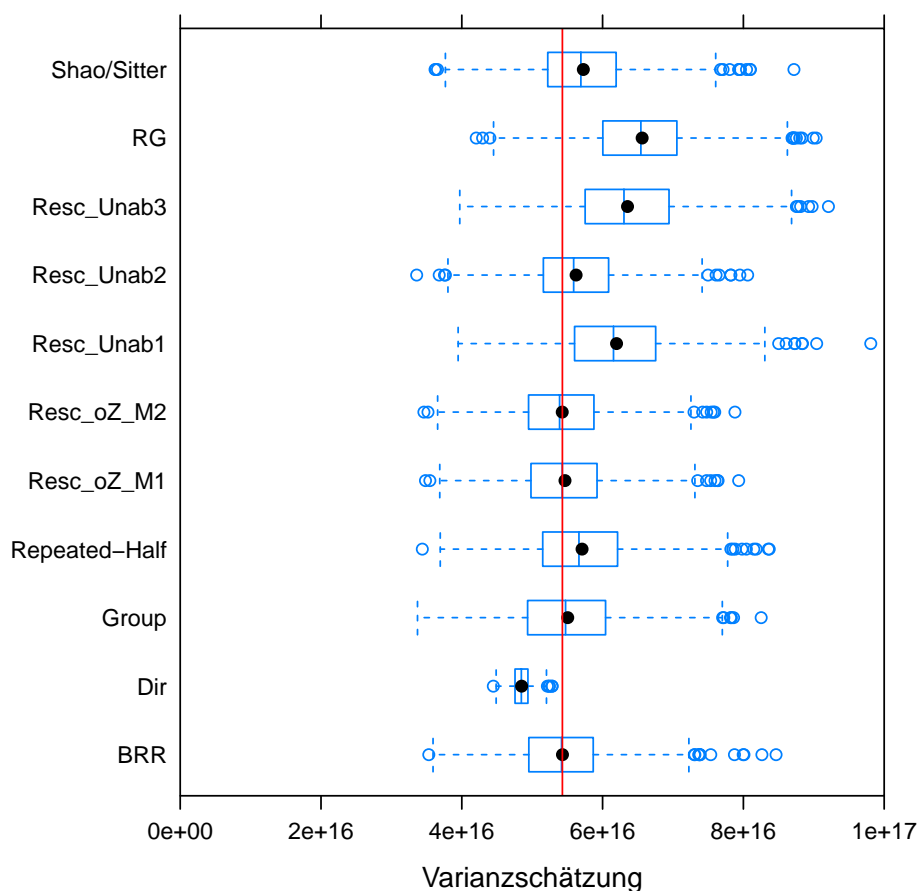


Abbildung 5.8: Vergleich der Varianzschätzer im Fall von Mittelwertimputation, kleiner Auswahlsätze, Entstehung variabler fehlender Werte in der Grundgesamtheit und MAR

An den Ergebnissen in Abbildung 5.8 ist zu sehen, dass insbesondere der BRR (*BRR*), der Delete-a-Group-Jackknife (*Group*) und die beiden Modifikationen des Rescaling-Bootstrap ohne Zurücklegen (*Resc_oZ_M1* bzw. *Resc_oZ_M2*) zu den besten und unverzerrten Schätzungen führen. Der BRR ist für einstufige stratifizierte Designs entwickelt worden und führt häufig für solche Designs zu guten Ergebnissen. Zwischen den beiden Modifikationen des

¹⁰Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.2 im Anhang A.

Rescaling-Bootstrap ohne Zurücklegen besteht kein großen Unterschied. Diese unterscheiden sich dahingehend, dass in der Modifikation 2 auch die Imputationswerte der nicht für die Substichprobe gezogenen Elemente mittels Mittelwertimputation reimputiert werden. Es liegen kleine Auswahlätze vor und, wie in Abschnitt 4.5.2 bzw. 4.5.4 beschrieben, besitzen die nicht gezogenen Elemente nur einen geringen Einfluss auf die Varianzschätzung, wodurch nur marginale Unterschiede bestehen. Auch der Delete-a-Group-Jackknife liefert gute Ergebnisse. Seine Umsetzung über Formel (4.17) ermöglicht eine gute Anpassung an das stratifizierte Design. Er ist jedoch nicht ganz so effizient wie der Rescaling-Bootstrap ohne Zurücklegen.

Der unabhängige Rescaling-Bootstrap führt in Abhängigkeit der Substichstichprobengröße zu drastischen Überschätzungen der Varianz. Dies gilt insbesondere für die Methode *Resc_Unab3*, bei der die nach Formel (4.22) von RAO und WU (1988) vorgeschlagene optimale Stichprobengröße verwendet wird. Formel (4.22) führt in diesem Fall unter den gegebenen Parameter zu Substichprobengrößen in jeder Schicht für welche $n_h^* > n_h$ gilt. In Abschnitt 4.5.1 wurde schon erwähnt, dass dies mit Problemen bei der Berechnung des Rescaling-Bootstrap einhergehen kann. Durch die von MASHREGHI et al. (2014) vorgeschlagene Wahl von $n_h - 3$ (Schätzer *Resc_Unab1*) kann die Verzerrung zwar etwas reduziert werden, die Überschätzung ist aber immer noch sehr groß. Erst durch eine Substichprobengröße von $n_h^* = n_h/2$ (Schätzer *Resc_Unab2*), wie sie im Rahmen des Rescaling-Bootstrap ohne Zurücklegen verwendet wird, kann die Verzerrung drastisch verringert werden. Komplette Beseitigung kann die Überschätzung auch bei dieser Größe nicht. Aber die Ergebnisse zeigen, dass die Wahl der Substichprobengröße der zentrale Parameter für das Gelingen des unabhängigen Rescaling-Bootstrap ist. Es ist jedoch auch zu beachten, dass hier eine wichtige Annahme des unabhängigen Rescaling-Bootstrap aus Abschnitt 4.7.2 nicht eingehalten werden konnte. Die Schichten entsprechen nicht den Imputationsklassen. Bei Einhaltung dieser Annahme können daher deutlich bessere Ergebnisse resultieren.

Auch der Shao und Sitter-Bootstrap (*Shao/Sitter*) und der Repeated-Half-Sample-Bootstrap (*Repeated-Half*) führen zu einer Überschätzung der Varianz. Dies steht im Zusammenhang mit der fehlenden Endlichkeitskorrektur beider Verfahren. Im Falle hoher Auswahlätze wären die Überschätzungen noch größer.

Die Random-Group-Methode (*RG*) liefert verzerrte Ergebnisse. Dies hängt mit dem Substichprobenbildungsprozess des Verfahrens zusammen. Diese Problematik wurde schon in Abschnitt 4.2 beschrieben. Die Substichproben stellen die einzelnen gebildeten Gruppen dar. Die Schätzungen in jeder Replikation basieren nur auf sehr wenigen Einheiten, welche entsprechend hochgewichtet werden. Auf diesem Wege sind die Varianzschätzer nicht besonders stabil. Des Weiteren verwendet die Methode keine Endlichkeitskorrektur, wodurch die Überschätzungen noch größer werden.

Der direkte Varianzschätzer (*Dir*) aus (3.43) führt in dieser Anwendung zu deutlichen Unterschätzungen. Dies liegt in Zusammenhang mit dem Ausfallmechanismus, welcher in diesem Szenario MAR ist. In Abschnitt 3.6.1 wurde beschrieben, dass dieser Varianzschätzer auf MCAR basiert. Im Fall dieses Ausfallmechanismus im Abschnitt 5.2.3.3 wird zu sehen sein, dass die Ergebnisse auf Basis des direkten Varianzschätzers aus (3.43) nahezu unverzerrt sind.

Die dargestellten Ergebnisse beziehen sich auf den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit (z_var_GG). Abbildung 5.9 beinhaltet den Vergleich zu den Fällen der Entstehung fehlender Werte in der Stichprobe (z_Stich) und eines fixen Response-Vektors (z_var_zfix) in der Grundgesamtheit für den Rescaling-Bootstrap ohne Zurücklegen. Da kleine Auswahlätze vorliegen, besitzt die Varianzkomponente bezüglich

fehlender Werte V_{NR}^1 einen sehr kleinen Anteil. Die Fälle der Entstehung variabler fehlender Werte in der Grundgesamtheit und eines fixen Response-Vektors unterscheiden sich daher kaum. Auch im Fall der Entstehung fehlender Werte in der Stichprobe bestehen kaum Unterschiede. Dieser Sachverhalt wird näher in Abschnitt 5.3.1 erläutert. Der Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit und der Entstehung fehlender Werte in der Stichprobe liefern sehr ähnliche Stichproben und Response-Vektoren. Dies hängt damit zusammen, dass fehlende Werte ausschließlich in Abhängigkeit der kategorialen Hilfsvariable erzeugt werden. Besonderheiten des Stichprobendesigns besitzen keinen Einfluss. Bezüglich der gesamten Varianz besteht kein Unterschied, sondern lediglich in der Höhe ihrer einzelnen Komponenten.

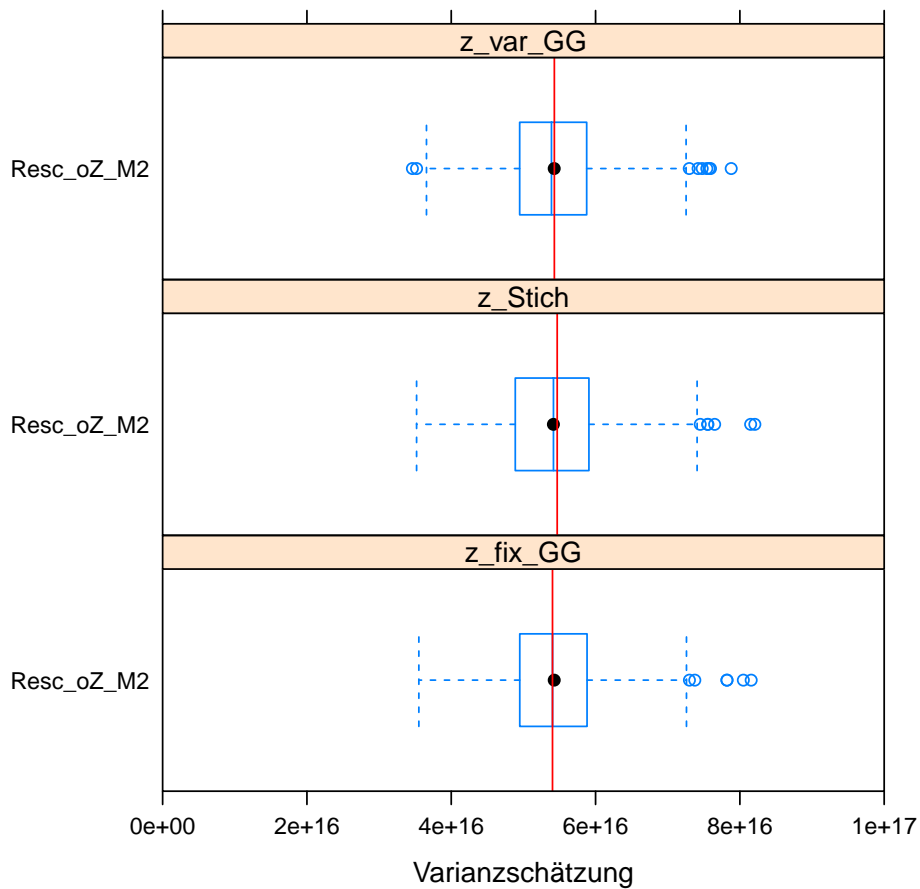


Abbildung 5.9: Vergleich des Rescaling-Bootstrap ohne Zurücklegen für die drei Fälle an fehlenden Werten bei kleinen Auswahlsätzen, MAR und Mittelwertimputation

5.2.1.2 Verhältnisimputation

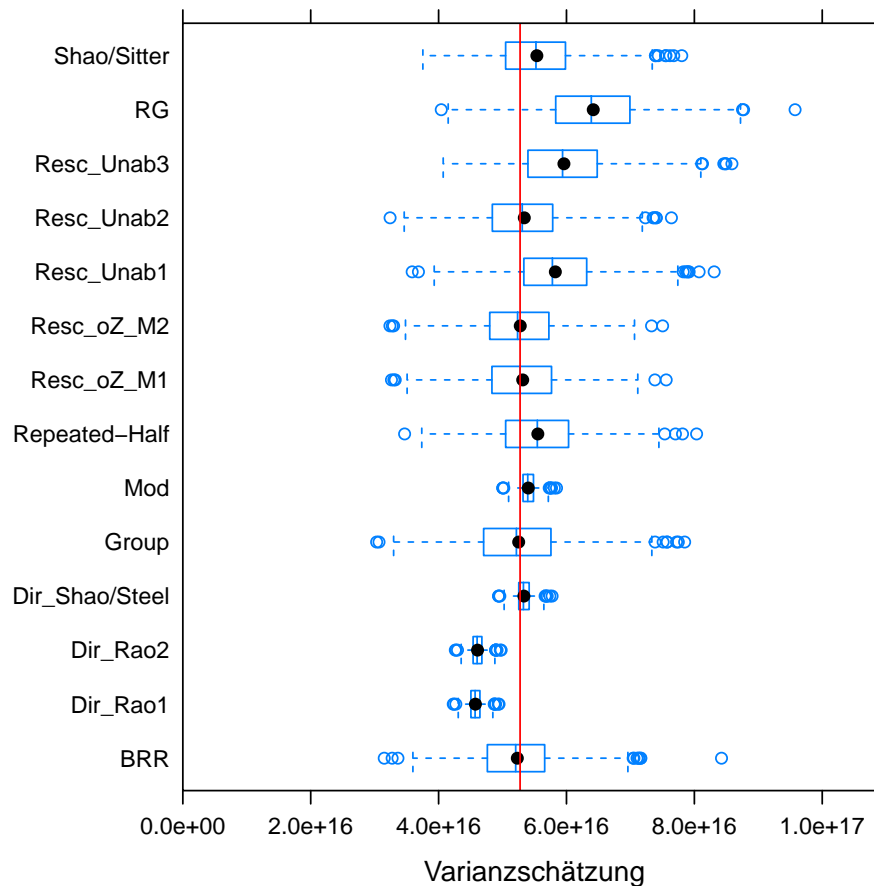


Abbildung 5.10: Vergleich der Varianzschätzer im Fall von Verhältnisimputation, kleiner Auswahlsätze, Entstehung variabler fehlender Werte in der Grundgesamtheit und MAR

Abbildung 5.10 zeigt die Varianzschätzergebnisse bei Verwendung von Verhältnisimputation und Erzeugung variabler fehlender Werte in der Grundgesamtheit mit Bezug auf die metrische Hilfsvariable aus Modell (5.1).¹¹ Ansonsten gelten die gleichen Parameterkonstellationen wie im Falle der Mittelwertimputation. Die Ergebnisse zwischen beiden Szenarien sind ähnlich. Der BRR, der Delete-a-Group-Jackknife (*Group*) und die beiden Modifikationen des Rescaling-Bootstrap ohne Zurücklegen (*Resc_oZ_M1* bzw. *Resc_oZ_M2*) führen wieder zu nahezu unverzerrten Schätzungen. Die Varianz der Varianz beim Delete-a-Group-Jackknife ist jedoch etwas höher. Der Repeated-Half-Sample-Bootstrap (*Repeated-Half*) und der Shao und Sitter-Bootstrap (*Shao/Sitter*) führen zu Überschätzungen aufgrund der fehlenden Endlichkeitskorrektur. Die Random-Group-Methode (*RG*) liefert wieder sehr instabile Schätzungen. Auffallend sind in diesem Szenario jedoch die guten Schätzungen beim unabhängigen Rescaling-Bootstrap. Dies gilt aber nur für die Variante 2 (*Resc_Unab2*), bei welcher die Substichprobe der Hälfte der ursprünglichen Stichprobengröße und damit derjenigen des Rescaling-Bootstrap ohne Zurücklegen entspricht. Die Verwendung von $n_h^* = n_h - 3$ (*Resc_Unab1*) bzw. der in

¹¹Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.3 im Anhang A.

(4.22) vorgeschlagenen Substichprobengröße (*Resc_Unab3*) führen wieder zu deutlichen Überschätzungen der Varianz. Es gelten hier aber auch weiterhin die Probleme der Annahmeverletzung der Übereinstimmung von Imputationsklasse und Schicht (vgl. Abschnitt 4.7.2).

Des Weiteren wurden bei der Verhältnisimputation noch die Varianzschätzer aus Abschnitt 3.6.3 berücksichtigt. Für alle diese Schätzer resultiert eine geringe Varianz der Varianz. Dabei weist der direkte Varianzschätzer *Dir_Shao/Steel* von SHAO und STEEL (1999) aus (3.49) die geringste Verzerrung auf. Unter der gegebenen Parameterkonstellation ist diese Methode im Falle der Verhältnisimputation vorzuziehen. Es bleibt jedoch anzumerken, dass seine Anwendbarkeit in dieser Form auf die Verhältnisimputation beschränkt ist. Wie bei allen direkten Varianzschätzer muss er für andere Imputationsmethoden separat hergeleitet werden, was unter Umständen nicht immer möglich ist. Bei der modellassistierenden Varianzschätzung (*Mod*) aus Formel (3.50) sind kleinere Überschätzungen zu beobachten. Das dem Verfahren zugrunde liegende Modell gibt unter Umständen das tatsächliche Imputationsmodell oder den Zusammenhang des Untersuchungsmerkmals y zur Hilfsvariable x nicht vollständig wieder.

Die direkten Varianzschätzer von RAO (1990) (*Rao1* bzw. *Rao2*) aus Formel (3.47) bzw. (3.48) sind sehr stark verzerrt, mit großen Unterschätzungen. Der Grund könnte hier in der mangelnden Vereinbarkeit dieser Methoden mit dem Stichprobendesign und dem Ausfallmechanismus dieses Szenarios liegen.

5.2.1.3 Regressionsimputation

Auch bei Regressionsimputation in Abbildung 5.11 resultieren ähnliche Ergebnisse.¹² Es zeigen sich die guten Schätzungen mittels Rescaling-Bootstrap ohne Zurücklegen (*Resc_oZ_M1* bzw. *Resc_oZ_M2*), sowie dem BRR und dem Delete-a-Group-Jackknife (*Group*), wobei die beiden letzten ein klein wenig unterschätzen. Große Unterschiede zeigen sich jedoch bezüglich des unabhängigen Rescaling-Bootstrap. Im Rahmen der Mittelwertimputation war es sinnvoll für die Substichprobengröße die Hälfte der ursprünglichen Stichprobengröße zu nehmen (vgl. Schätzer *Resc_Unab2*). Bei der Regressionsimputation ist die von RAO und WU (1988) vorgeschlagene Substichprobengröße aus (4.22) die bessere Wahl (vgl. Schätzer *Resc_Unab3*). Der Vergleich mit den Ergebnissen der Mittelwertimputation zeigt, wie schwierig es ist, die optimale Substichprobengröße beim unabhängigen Rescaling-Bootstrap zu bestimmen. Sie ist von einigen Faktoren abhängig, wie in diesem Fall zum Beispiel von der Imputationsmethode. Diese Problematik erschwert seine Anwendung. Eine Überprüfung der Eignung der direkten Varianzschätzer aus Abschnitt 3.6.4 wird in Abschnitt A.3.2 im Anhang A vorgenommen. Bezüglich dieser Schätzer sind andere Parameterkonstellationen von Interesse. Daher werden sie von den Resampling-Verfahren getrennt untersucht.

¹²In diesem Szenario gelten die Parameterkonstellationen der Verhältnisimputation aus Abschnitt A.2.2.3 im Anhang A.

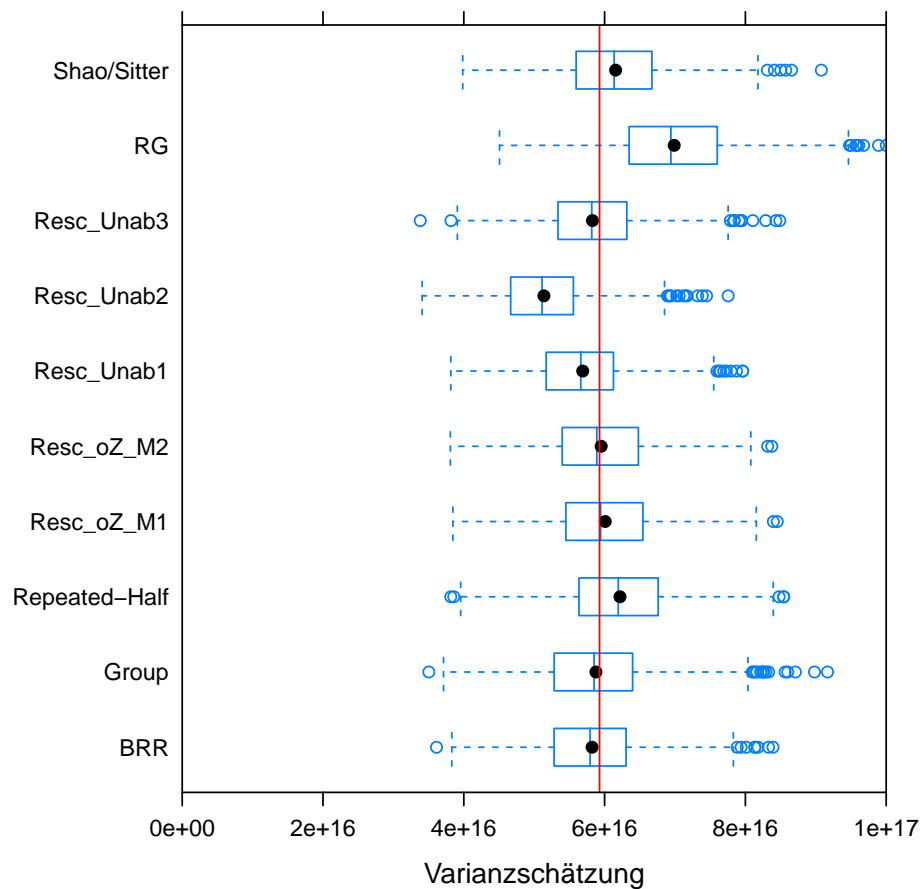


Abbildung 5.11: Vergleich der Varianzschätzer im Fall von Regressionsimputation, kleiner Auswahlsätze, Entstehung variabler fehlender Werte in der Grundgesamtheit und MAR

5.2.1.4 Hot-Deck-Random-Imputation

Das erste zufällige Imputationsverfahren dieser Simulationsstudie ist die Hot-Deck-Random-Imputation.¹³ Die Varianzschätzungen für diese Methode im Falle kleiner Auswahlsätze, Erzeugung fehlender Werte in der Grundgesamtheit sowie des Vorliegens von MAR sind in Abbildung 5.12 angegeben. Die Resultate in Bezug auf den Vergleich der einzelnen Methoden sind den Ergebnissen des deterministischen Falles, insbesondere für Mittelwertimputation bei der entsprechenden Parameterkonstellation, sehr ähnlich. Insbesondere die Parameter der Erzeugung fehlender Werte sind hier identisch. Fehlende Werte werden in Abhängigkeit der kategorialen Variablen Haushaltsgröße gebildet. Dementsprechend führen der BRR, der Delete-a-Group-Jackknife (*Group*) und die beiden Modifikationen des Rescaling-Bootstrap ohne Zurücklegen (*Resc_oZ_M1* bzw. *Resc_oZ_M2*) wiederum zu unverzerrten Schätzungen der Varianz. Die Random-Group-Methode (*RG*), der Shao und Sitter-Bootstrap (*Shao/Sitter*) und der Repeated-Half-Sample-Bootstrap (*Repeated-Half*) überschätzen aus den bekannten Gründen die Varianz. Der direkte Varianzschätzer (*Dir*) aus Formel (3.46) stellt im Vergleich zum direkten Varianzschätzer der Mittelwertimputation aus (3.43) nur eine Erweiterung um die zu schätzende Imputationskomponente dar.

¹³In diesem Szenario gelten die Parameterkonstellationen der Mittelwertimputation aus Abschnitt A.2.2.2 im Anhang A.

Im Falle von MAR zeigten sich schon im deterministischen Fall große Unterschätzungen. Und so verhält es sich auch für die Hot-Deck-Random-Imputation.

Für den unabhängigen Rescaling-Bootstrap zeigt sich im Fall der zufälligen Imputationsmethode jedoch eine Besonderheit. Die Ergebnisse liegen unabhängig von der Wahl der Substichprobengröße sehr dicht aneinander, unterschätzen die Varianz aber deutlich. Diese Resampling-Methode könnte in diesem Fall nicht mit der Adjustierung der imputierten Werte in jeder Substichprobe kompatibel sein, wodurch sich die Verzerrungen ergeben. Auch die beschriebene Verletzung der Annahme aus Abschnitt 4.7.2 bezüglich der von den Schichten abweichenden Imputationsklassen kann hier starke Auswirkungen haben. Hier bedarf es weiterer Untersuchungen.

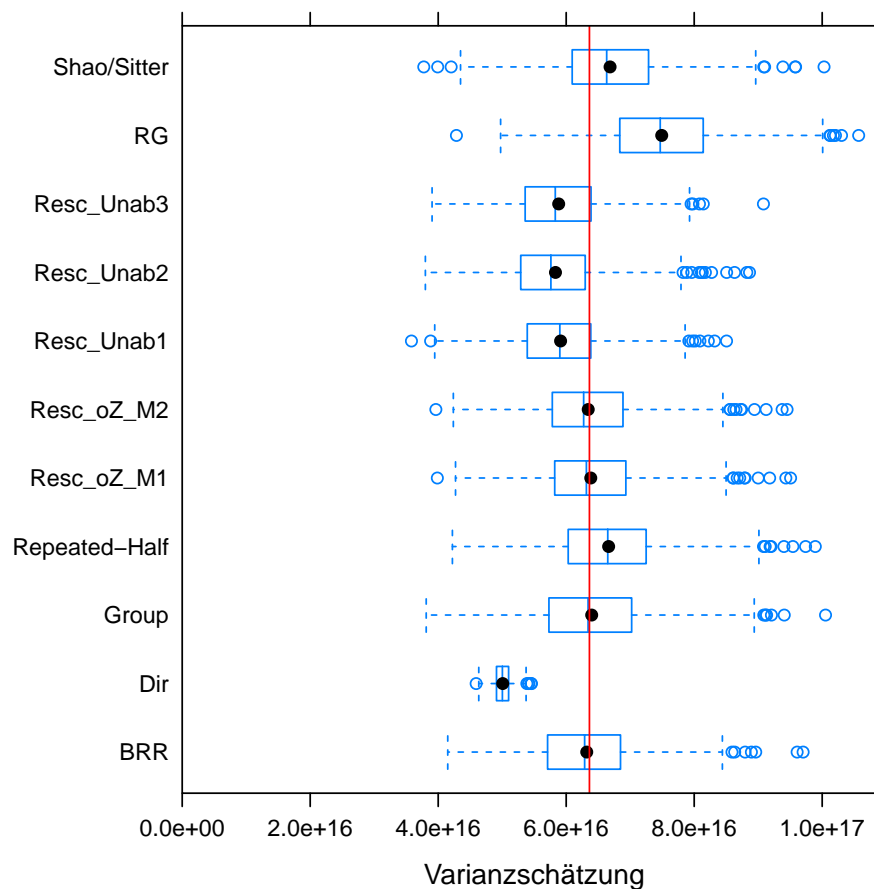


Abbildung 5.12: Vergleich der Varianzschätzer im Fall von Hot-Deck-Random-Imputation, kleiner Auswahlsätze, Entstehung variabler fehlender Werte in der Grundgesamtheit und MAR

Diese Ähnlichkeit der Ergebnisse zwischen Mittelwert- und Hot-Deck-Random-Imputation bezieht sich nur auf den ordinalen Vergleich der Schätzmethoden. Zwischen der Höhe der absoluten Varianzen und der Resampling-Verfahren bestehen große Unterschiede zwischen den beiden Imputationsmethoden. Die Hot-Deck-Random-Imputation ist im Gegensatz zur Mittelwertimputation ein zufälliges Imputationsverfahren. Die Varianz ist hier viel höher, da hier noch zusätzlich die Imputationskomponente geschätzt werden muss. Außerdem wird statt der Reimputation die Adjustierung der imputierten Werte innerhalb der Resampling-Verfahren angewandt, um die Imputation zu berücksichtigen. Dieser Sachver-

halt wurde in Abschnitt 4.1.2 erläutert. Wird bei zufälligen Imputationsverfahren wie Hot-Deck-Random-Imputation die Reimputation verwendet, können große Überschätzungen der Varianz entstehen. Dies wird in Abbildung 5.13 anhand des Rescaling-Bootstrap ohne Zurücklegen Modifikation 2 dargestellt. In Abbildung 5.12 wurde dieses Verfahren (*Resc_oZ_M2*) auf Basis der Adjustierung der imputierten Werte in jeder Substichprobe verwendet. Dieser Schätzer wird zur Abgrenzung gegenüber der Reimputation in Abbildung 5.13 mit *Resc_oZ_M2_Adjust* bezeichnet. Der obere Schätzer (*Resc_oZ_M2_Reimp*) greift hingegen auf die Reimputation zurück, also der einfachen Wiederholung der Hot-Deck-Random-Imputation der ursprünglichen Stichprobe in jeder Substichprobe. Es zeigt sich eine deutliche Überschätzung der Varianz unter der Verwendung der Reimputation. Die Adjustierung der imputierten Werte führt hingegen zu einer unverzerrten Schätzung. Der Grund liegt wie in Abschnitt 4.1.2 beschrieben darin, dass bei Resampling-Methoden unter Anwendung zufälliger Imputationsverfahren und unabhängiger Verwendung von Zufallstermen von Substichprobe zu Substichprobe Überschätzungen der Varianz auftreten, wenn die Substichprobengröße nicht der Größe der ursprünglichen Stichprobe entspricht. Und dies ist beim Rescaling-Bootstrap ohne Zurücklegen der Fall, da hier in jeder Substichprobe nur die Hälfte der zugrunde liegenden Stichprobe genommen wird.

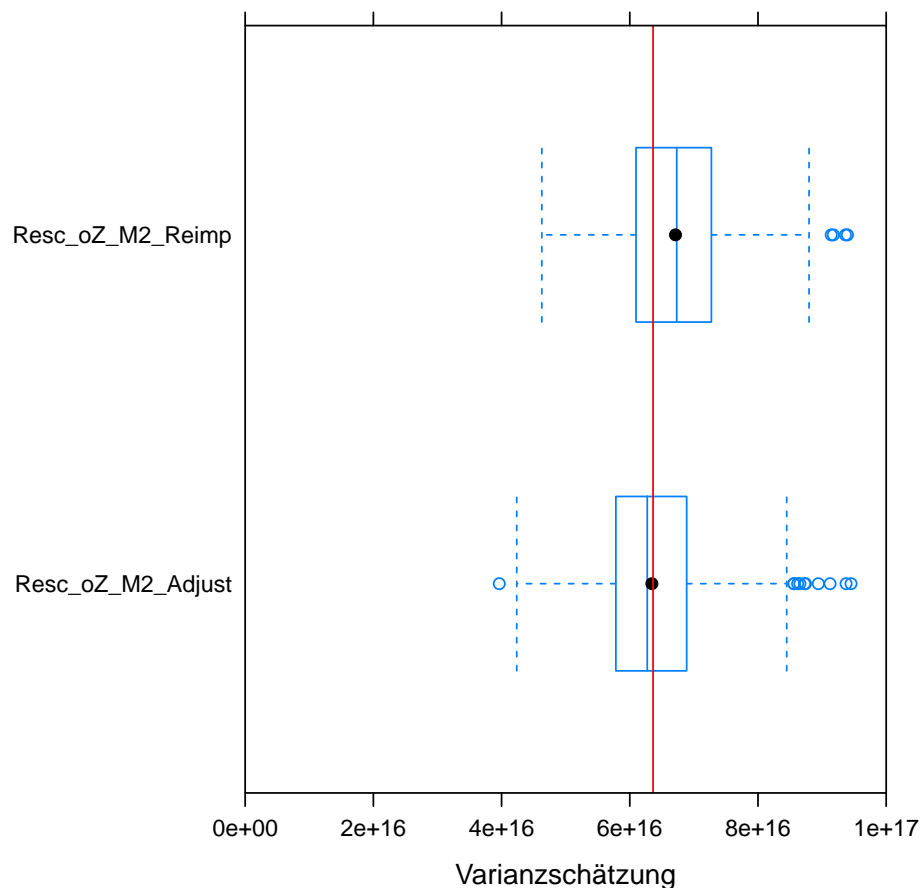


Abbildung 5.13: Vergleich der Reimputation und Adjustierung für den Rescaling-Bootstrap ohne Zurücklegen im Rahmen von Hot-Deck-Random-Imputation

5.2.1.5 Fazit bezüglich des Vergleichs verschiedener Varianzschätzmethoden

Die Varianzschätzungen mittels des BRR, die Modifikationen des Rescaling-Bootstrap und der Delete-a-Group-Jackknife führten in den verschiedenen Szenarien zu den besten Ergebnissen. Für letztere Methode war die Varianz der Varianz jedoch meist etwas höher. Bei einige Methoden zeigen sich Überschätzungen aufgrund der Nichtberücksichtigung einer Endlichkeitskorrektur und dem verwendeten Ziehungsmodell ohne Zurücklegen der ursprünglichen Stichprobe. Im Falle von Verhältnisimputation führte der direkte Varianzschätzer von SHAO und STEEL (1999) zu guten Ergebnissen. Für andere Imputationsmethoden konnten die direkten Varianzschätzer nicht überzeugen. Ein möglicher Grund lag im MAR-Ausfallmechanismus, wodurch die Annahmen dieser Methoden verletzt sind. Auch beim unabhängigen Rescaling-Bootstrap können Annahmeverletzungen zu den teilweise negativen Resultaten beigetragen haben. Hier stellte sich zusätzlich das Problem der Substichprobengrößenbestimmung.

5.2.2 Varianzschätzung für komplexere Imputationsmethoden

Im diesem Kapitel werden ausgewählte Resampling-Methoden bezüglich ihrer Eignung zur Schätzung der Varianz bei Verwendung komplexerer Imputationsverfahren überprüft. Es werden nicht mehr alle in Abschnitt 5.2.1 berücksichtigten Varianzschätzer verwendet. Zum einen sind die Laufzeiten der einzelnen Verfahren bei komplexeren Imputationsmethoden wie Nearest-Neighbour-Imputation oder Predictive-Mean-Matching deutlich länger. Zum anderen steht hier im Gegensatz zum vorherigen Abschnitt nicht der Methodenvergleich im Vordergrund. Das Ziel besteht in diesem Abschnitt in der Darstellung bestimmter Problematiken bei der Durchführung der Varianzschätzung bei komplexen Imputationsmethoden.

Im Rahmen dieses Abschnittes werden daher der BRR, der Rescaling-Bootstrap ohne Zurücklegen Modifikation 2,¹⁴ sowie der Shao und Sitter-Bootstrap berücksichtigt. Die Wahl der beiden ersten Verfahren wurde aufgrund ihrer guten Ergebnisse beim Methodenvergleich im Abschnitt 5.2.1 getroffen. Gegenüber des Delete-a-Group-Jackknife waren die Varianzen der Varianzen meist etwas geringer. Der Shao und Sitter-Bootstrap wird zur Kontrolle miteinbezogen, da zumindest ein Verfahren noch berücksichtigt werden soll, welches im Rahmen der Simulation in Abschnitt 5.2.1 nicht gut funktioniert hat. Bezüglich dieses Verfahrens zeigten sich Überschätzungen der Varianz aufgrund der fehlenden Integration der Endlichkeitskorrektur. Dieser negative Effekt ist jedoch gut interpretierbar. Daher wird das Verfahren hier verwendet.

5.2.2.1 Stochastische Regressionsimputation

Die erste vorgestellte komplexere Imputationsmethode ist die stochastische Regressionsimputation, da Imputationswerte aus einem deterministischen und einem zufälligen Anteil bestehen.¹⁵ Die Varianzschätzungen werden in Abbildung 5.14 gezeigt. Hier führt die Adjustierung der imputierten Werte in jeder Substichprobe mit den Erwartungswerten aus Tabelle 4.1 beim BRR, den Rescaling-Bootstrap ohne Zurücklegen Modifikation 2 zu fast unverzerrten Schätzungen.

¹⁴Der unabhängige Rescaling-Bootstrap wird in den nächsten Szenarien nicht mehr berücksichtigt. Daher ist im Folgenden der Rescaling-Bootstrap ohne Zurücklegen gemeint, wenn vom Rescaling-Bootstrap die Rede ist.

¹⁵In diesem Szenario gelten die Parameterkonstellationen der Verhältnisimputation aus Abschnitt A.2.2.3 im Anhang A.

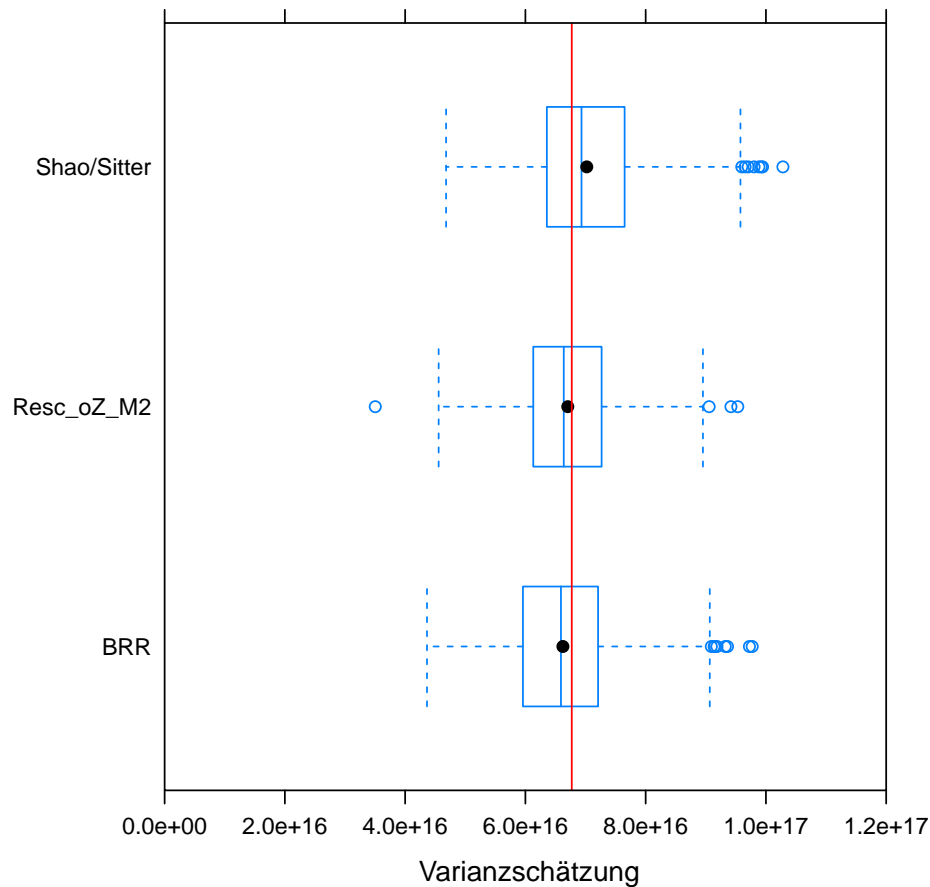


Abbildung 5.14: Vergleich der Varianzschätzer im Fall der stochastischen Regressionsimulation, der Entstehung variabler fehlender Werte in der Grundgesamtheit, kleiner Auswahlätze und MAR

5.2.2.2 Nearest-Neighbour-Imputation

Bei Nearest-Neighbour-Imputationen und ihrem Spezialfall Predictive-Mean-Matching resultiert die Komplexität bei der Anwendung von Resampling-Verfahren insbesondere aus der Bestimmung des nächsten Nachbarn in der Substichprobe. So zeigt sich bei der Nearest-Neighbour-Imputation in Abbildung 5.15 ein etwas kurioses Bild.¹⁶ Gezeichneten sich die Varianzschätzer in den Abschnitten 5.2.1 und 5.2.2.1 durch eine eher symmetrische Verteilung aus, liegt hier ein großer Anteil der Schätzungen unterhalb der roten Linie und des Benchmark. Auf der anderen Seite zeigen sich für alle drei Verfahren viele Ausreißer nach oben. Diese sind teilweise so hoch, dass sie nicht mehr im sichtbaren Bereich der Abbildung liegen. Diese unsymmetrischen Verteilungen können insbesondere aus den Reimputationen und damit der Bestimmung des nächsten Nachbarn in der Substichprobe resultieren. So kann zum Beispiel der nächste Nachbar einer fehlenden Einheit nicht für die Substichprobe gezogen werden. Dann ist es möglich, dass der Beobachtungswert einer Einheit, welche basierend auf den Hilfsvariablen die geringste Distanz zu der Einheit mit fehlendem Wert aufweist, stark vom ursprünglichen Imputationswert abweichen kann. Dies gilt insbesondere bei einer schiefen Verteilung des Untersuchungsmerkmals bei dem

¹⁶Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.4 im Anhang A.

Ausreißer als nächster Nachbar in Erscheinung treten können. Ist dies für mehrere fehlende Einheiten einer Substichprobe der Fall, kann die darauf basierende Varianzschätzung mittels Resampling-Verfahren deutlich zu hoch ausfallen, da der Punktschätzer dieser Substichprobe von denjenigen der übrigen Substichproben stark abweichen kann. Aus diesem Vorgang kann die stark asymmetrische Verteilung resultieren. Die Mittelwerte der Varianzschätzungen liegen jedoch zumindest für den BRR und Rescaling-Bootstrap ohne Zurücklegen sehr nahe am Benchmark. Beim Rescaling-Bootstrap ohne Zurücklegen führen die Ausreißer zu kleineren Überschätzungen der Varianz. Der BRR führt hier zu den besten Ergebnissen, da diese Methode durch die ausbalancierte Auswahl über die Hadamard-Matrizen weniger ausreißeranfällig ist. Der Substichprobenbildungsprozess in Verbindung mit der Reimputation ist hier entscheidend. Dieser kann auch für die großen Unterschätzungen beim Shao und Sitter-Bootstrap verantwortlich sein. Hier entsteht durch das Substichprobenziehungsschema mit Zurücklegen und die darauf basierende Reimputation zu wenig Variabilität. Hier wären Überschätzungen aufgrund der fehlenden Endlichkeitskorrektur erwartet worden. Durch die Ziehung mit Zurücklegen ist es denkbar, dass für Einheiten mit fehlendem Wert bestimmte Spender über die Substichproben hinweg zu häufig als nächster Nachbar klassifiziert werden und dadurch zu wenig Variation entsteht. Dieser Sachverhalt bedarf weiterführender Untersuchungen.

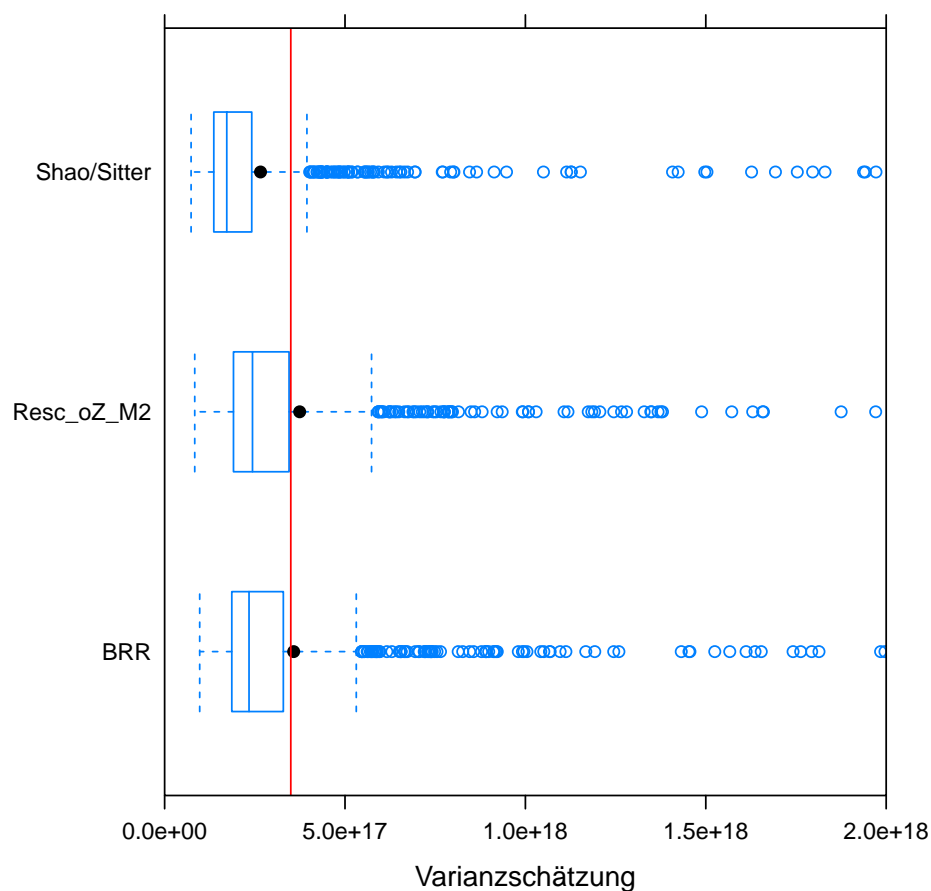


Abbildung 5.15: Vergleich der Varianzschätzer im Fall von Nearest-Neighbour-Imputation, der Entstehung variabler fehlender Werte in der Grundgesamtheit, kleiner Auswahlätze und MAR

Die Entstehung der Ausreißer wird natürlich sehr stark von den verwendeten Variablen begünstigt. Das Einkommen als Untersuchungsmerkmal ist sehr schief verteilt. Die zur Distanzberechnung verwendete metrische Hilfsvariable wurde nach dem in (5.1) beschriebenen Modell erzeugt und besitzt zum Untersuchungsmerkmal nur eine geringe Korrelation. Um diesen Sachverhalt näher zu untersuchen, wird die Varianzschätzung durch die drei Resampling-Methoden noch für ein weiteres Untersuchungsmerkmal durchgeführt.¹⁷ In Abbildung 5.16 wird die Gesamtanzahl der Kategorie 2, der Variable Hausgröße aus Tabelle 5.2 geschätzt. Als Hilfsvariable wird das bisherige Untersuchungsmerkmal Einkommen verwendet. Für diesen Fall ist in Abbildung 5.16 zu sehen, dass sich für den Rescaling-Bootstrap ohne Zurücklegen und den BRR unverzerrte Resultate ergeben. Die Verteilung der Schätzergebnisse ist deutlich symmetrischer und die Ausreißeranzahl geringer. Weitere detaillierte Vergleiche zum vorherigen Szenario, insbesondere bezüglich der absoluten Höhe der Ausreißer können nicht durchgeführt werden, da Skalierungen und absolute Höhen der Varianzen stark unterschiedlich sind.

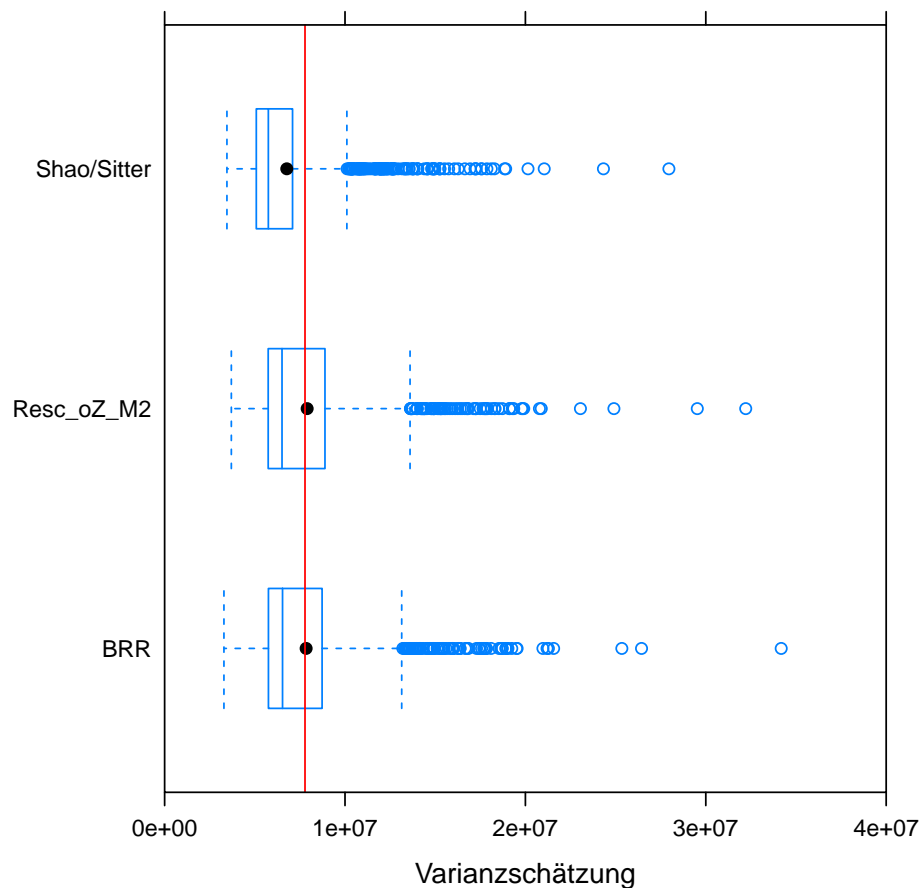


Abbildung 5.16: Vergleich der Varianzschätzer im Fall von Nearest-Neighbour-Imputation bei Schätzung der Gesamtanzahl der Kategorie 2 der Haushaltsgröße und des Einkommens als Hilfsvariable

¹⁷Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.4 im Anhang A.

5.2.2.3 Predictive-Mean-Matching

Die Varianzschätzer für eine zweite Variante von Nearest-Neighbour-Verfahren, dem Predictive-Mean-Matching, werden in Abbildung 5.17 gegeben.¹⁸

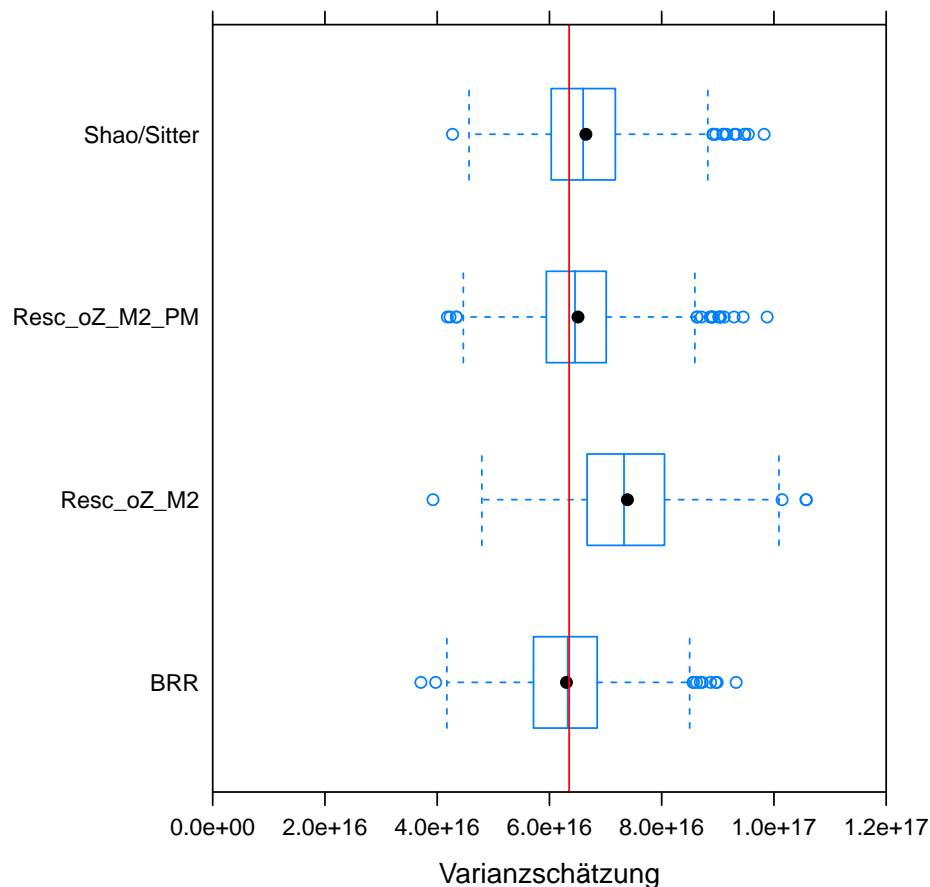


Abbildung 5.17: Varianzschätzung im Fall von Predictive-Mean-Matching, der Entstehung variabler fehlender Werte in der Grundgesamtheit, kleiner Auswahlätze und MAR

Wie in Abschnitt 5.1.6 beschrieben wird das Imputationsverfahren innerhalb der Schichten durchgeführt, da eine Anwendung auf die gesamte Stichprobe rechentechnisch zu aufwendig wäre. Die kategoriale Hilfsvariable Haushaltstyp, welche sonst zur Bildung der Imputationsklassen verwendet wird, ist schon neben der metrischen Variable im Imputationsmodell berücksichtigt. Zunächst werden die Schätzungen des Rescaling-Bootstrap ohne Zurücklegen (*Resc_oZ_M2*) und des BRR betrachtet. Hier lässt sich feststellen, dass der BRR die Varianz unverzerrt schätzt. Der Rescaling-Bootstrap hingegen überschätzt die Varianz deutlich. Der Grund hierfür könnte in der Schätzung der für das Predictive-Mean-Matching notwendigen prognostizierten Werten¹⁹ liegen in Verbindung mit dem Substichprobenbildungsprozess. Über die Zufallsauswahl kann es beim Rescaling-Bootstrap innerhalb der Schichten zu einer unglücklichen Einteilung von gezogenen und nicht gezogen

¹⁸Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.5 im Anhang A.

¹⁹Vgl. hierzu Abschnitt 2.2.7.

Elementen kommen, innerhalb derer die Reimputation stattfindet. Dadurch kann es zu einer schlechten Schätzung des zugrunde liegenden Regressionsmodells und einer suboptimalen Bestimmung des nächsten Nachbarn kommen. Die ausbalancierte Auswahl beim BRR wirkt sich hier vorteilhaft aus. Für den Rescaling-Bootstrap müsste daher eine stabilere Schätzung des zugrunde liegenden Regressionsmodells realisiert werden. Daher wird in dieser Arbeit vorgeschlagen, die prognostizierten Werte vor der Substichprobenziehung auf Basis der gesamten Stichprobe zu schätzen. Die auf diesem Wege erhaltenen prognostizierten Werte werden bei der Reimputation innerhalb der Schichten und separat zwischen den für die Substichprobe gezogenen und nicht gezogenen Einheiten für Bestimmung des nächsten Nachbarn verwendet. Dies geschieht beim Schätzer *Resc_oZ_M2_PM*. Dadurch resultiert eine fast unverzerrte Varianzschätzung. Um eine stabile Varianzschätzung mittels des Rescaling-Bootstrap im Falle von Predictive-Mean-Matching zu gewährleisten, sollte die Schätzung der prognostizierten Werte daher schon vor der Substichprobenziehung erfolgen.

5.2.2.4 Fractional-Imputation

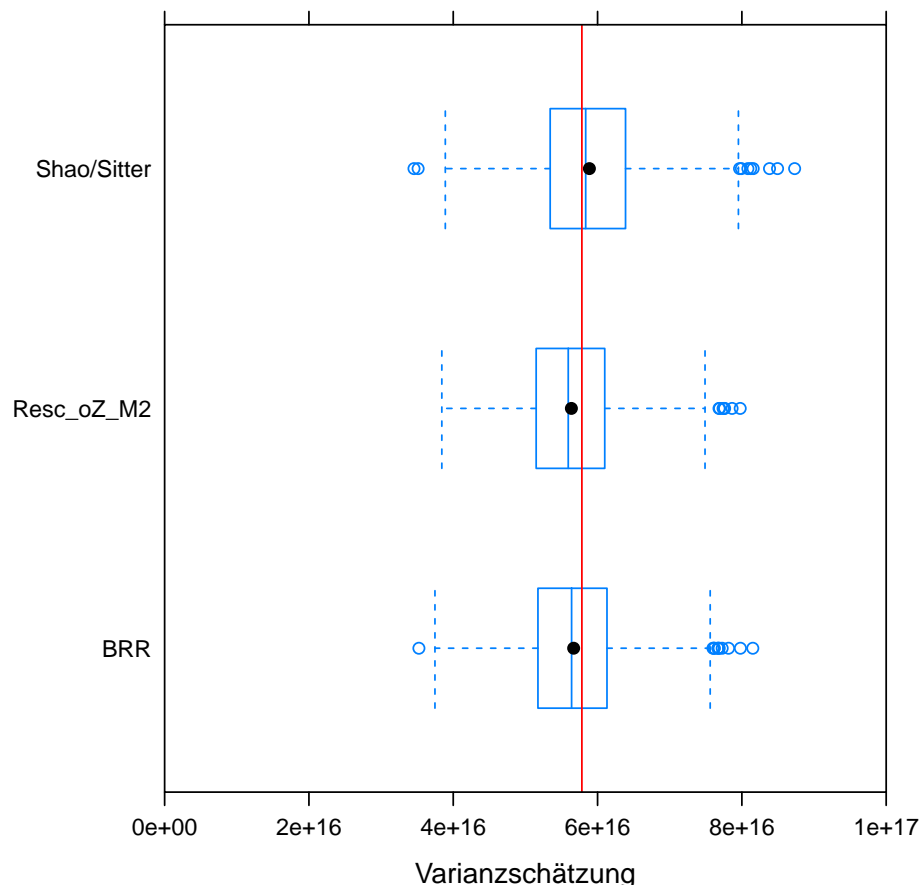


Abbildung 5.18: Vergleich der Varianzschätzer im Fall von Fractional-Imputation, der Entstehung variabler fehlender Werte in der Grundgesamtheit, kleiner Auswahlätze und MAR

In diesem Abschnitt wird zum ersten Mal eine Methode berücksichtigt, bei welcher sich der Imputationswert aus mehreren Einzelwerten berechnet.²⁰ Abbildung 5.18 zeigt die Ergebnisse der Varianzschätzung bei Verwendung von Fractional-Imputation. Es sind kleinere Ungenauigkeiten zu erkennen. Dies kann insbesondere mit der Komplexität der Imputationsmethode in Verbindung stehen. Wie in Abschnitt 2.2.9 beschrieben worden ist, liegt hier formal eine Hot-Deck-Random-Imputation mit D verschiedenen Imputationswerten vor. Jedoch erfolgt über die Durchschnittsbildung eine Annäherung an die Verteilungen deterministischer Imputationsverfahren.

5.2.2.5 Fazit bei Anwendung der Resampling-Methoden für komplexe Imputationsmethoden

Interessant waren hier besonders die Ergebnisse für die Verfahren, welche auf den nächsten Nachbar als Imputationswert abstellen. Herausforderungen bezüglich der Varianzschätzung sind insbesondere durch die Auswahl des nächsten Nachbarn in der Substichprobe gegeben. Hier zeigte der BRR durch seine ausbalancierte Auswahl Vorteile. Beim Rescaling-Bootstrap ohne Zurücklegen ist eine weitere Modifikation im Rahmen von Predictive-Mean-Matching notwendig. Hier müssen die prognostizierten Werte auf Basis des zugrunde liegenden Regressionsmodells vor der Substichprobenziehung bestimmt werden.

5.2.3 Untersuchungen zu bestimmten Problematiken der Varianzschätzung unter Imputation

Dieser Abschnitt der Simulationsergebnisse dient der Darstellungen bestimmter Problematiken der Varianzschätzung unter Imputation, wie sie in der Praxis auftreten können. Wie im vorherigen Abschnitt steht hier der Methodenvergleich nicht mehr im Vordergrund. Es wird wieder eine Konzentration auf die drei Resampling-Verfahren Rescaling-Bootstrap ohne Zurücklegen, den BRR und den Shao und Sitter-Bootstrap stattfinden. Im Abschnitt 5.2.3.3 beim Vergleich verschiedener Antwortmechanismen bei der Mittelwertimputation wird zusätzlich der direkte Varianzschätzer für Mittelwertimputation einbezogen werden. Grund hierfür ist die Überprüfung der in Abschnitten 3.6 und 5.2.1.1 aufgestellten Behauptung seiner Unverzerrtheit im Falle von MCAR.

5.2.3.1 Erhöhung der Rate der fehlenden Werte

Abbildung 5.19 zeigt eine Erhöhung der Rate der fehlenden Werten auf ungefähr 46 %.²¹ Hierzu wurde der Parameter $\tilde{\alpha}$ in Formel (5.6) heruntersgesetzt, so dass weniger Einheiten teilnehmen. Es ist zu sehen, dass der Rescaling-Bootstrap ohne Zurücklegen und der BRR, beide ohne Berücksichtigung von Korrekturfaktoren für die Varianzkomponente bezüglich fehlender Werte, die Varianz etwas unterschätzen. Der Shao und Sitter-Bootstrap liegt am dichtesten am Benchmark. Hier gilt es jedoch zu beachten, dass zwei gegenteilige Effekte auftreten. Wie der Rescaling-Bootstrap und der BRR würde diese Methode die Varianz unterschätzen. Aber der Shao und Sitter-Bootstrap verwendet keine Endlichkeitskorrektur, durch welche Überschätzungen resultieren, welche die Unterschätzungen gerade ausgleichen. Damit ist das Verfahren nicht unbedingt unverzerrt, sondern profitiert von diesen beiden glücklichen Umständen.

²⁰In diesem Szenario gelten die Parameterkonstellationen der Mittelwertimputation aus Abschnitt A.2.2.2 im Anhang A.

²¹Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.6 im Anhang A.

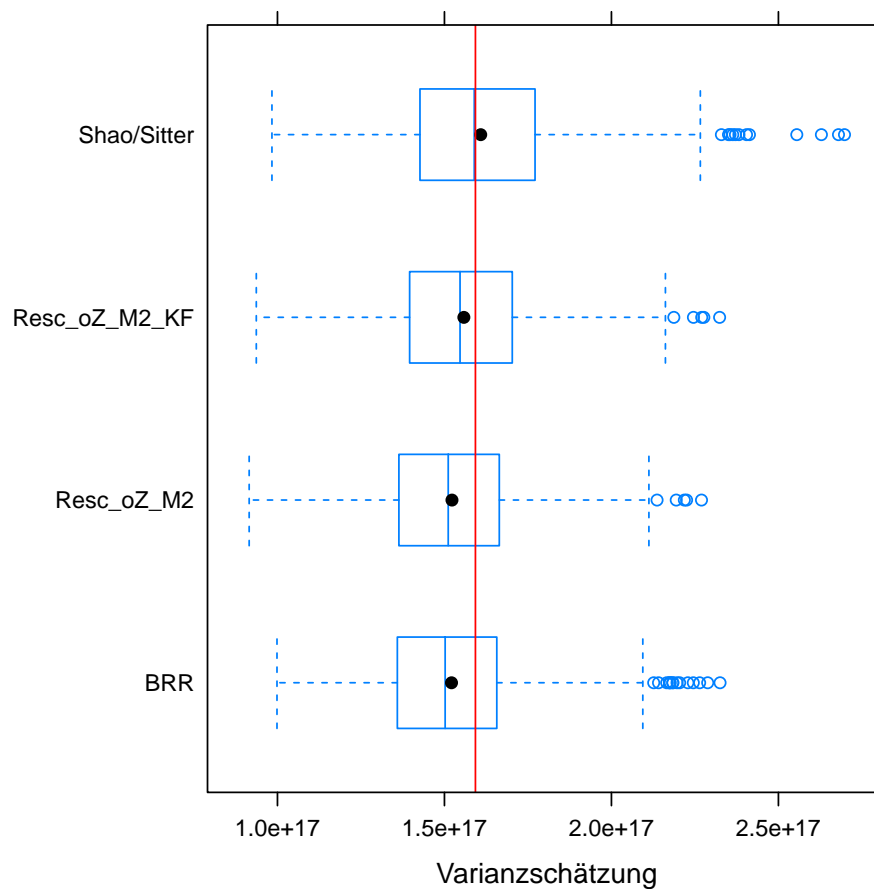


Abbildung 5.19: Erhöhung der Ausfallrate bei Mittelwertimputation, Erzeugung variabler fehlender Werte in der Grundgesamtheit, kleinen Auswahlsätze und MAR

Zentral in diesem Szenario sind damit die durch die Erhöhung der Rate der fehlenden Werte gegebenen Verzerrungen. Diese lassen sich zum einen durch den fehlenden Korrekturfaktor erklären. Wie in Abschnitt 3.4.3 erläutert, steigt die Varianzkomponente bezüglich fehlender Werte *ceteris paribus* bei Erhöhung der Ausfallrate. Diese Varianzkomponente kann so groß werden, dass die Berücksichtigung der Korrekturfaktoren erforderlich wird. Daher wird in Abbildung 5.19 noch der Rescaling-Bootstrap ohne Zurücklegen angegeben, welcher den in Abschnitt 4.7.3 beschriebenen Korrekturfaktor für Mittelwertimputation beinhaltet (*Resc_oZ_M2_KF*).²² Aber auch diese Methode bringt noch kleinere Unterschätzungen mit sich. Dies kann mit der größeren Ausfallrate erklärt werden. Mit einer größeren Anzahl fehlender Werte wird auch die Varianzschätzung instabiler. Dadurch können solche Ungenauigkeiten resultieren.

5.2.3.2 Erzeugung fehlender Werte in den gering besetzten Kategorien

Nach Tabelle 5.2 sind die Kategorien 2 und 3 der Hilfsvariable Haushaltstyp zur Bildung der Imputationsklassen am geringsten besetzt. Werden fehlende Werte bezüglich

²²Die Simulationsergebnisse für Resampling-Verfahren, welche diese Korrekturfaktoren zur Abdeckung der Komponente bezüglich fehlender Werte berücksichtigen, werden ausführlich in Abschnitt 5.2.4 gegeben.

des Untersuchungsmerkmals hauptsächlich in diesen Kategorien erzeugt,²³ ist in Abbildung 5.20 zu sehen, dass die Varianzschätzungen mit Ungenauigkeiten in Form kleinerer Unterschätzungen einhergehen. Da fehlende Werte schwerpunktmäßig in Kategorien erzeugt werden, welche im Vergleich zu den anderen Kategorien deutlich schwächer besetzt sind, treten kleinere Instabilitäten bei der Schätzung auf. In den übrigen Kategorien sind jedoch noch genug Einheiten vorhanden, so dass diese Ungenauigkeiten nicht drastisch sind. Dies gilt für den BRR und den Rescaling-Bootstrap. Beim Shao und Sitter-Bootstrap sind weiterhin Überschätzungen zu erkennen, welche jedoch nicht so groß sind. Dies liegt daran, dass diese Ungenauigkeiten erzeugten Unterschätzungen einen gegenteiligen Effekt ausüben.

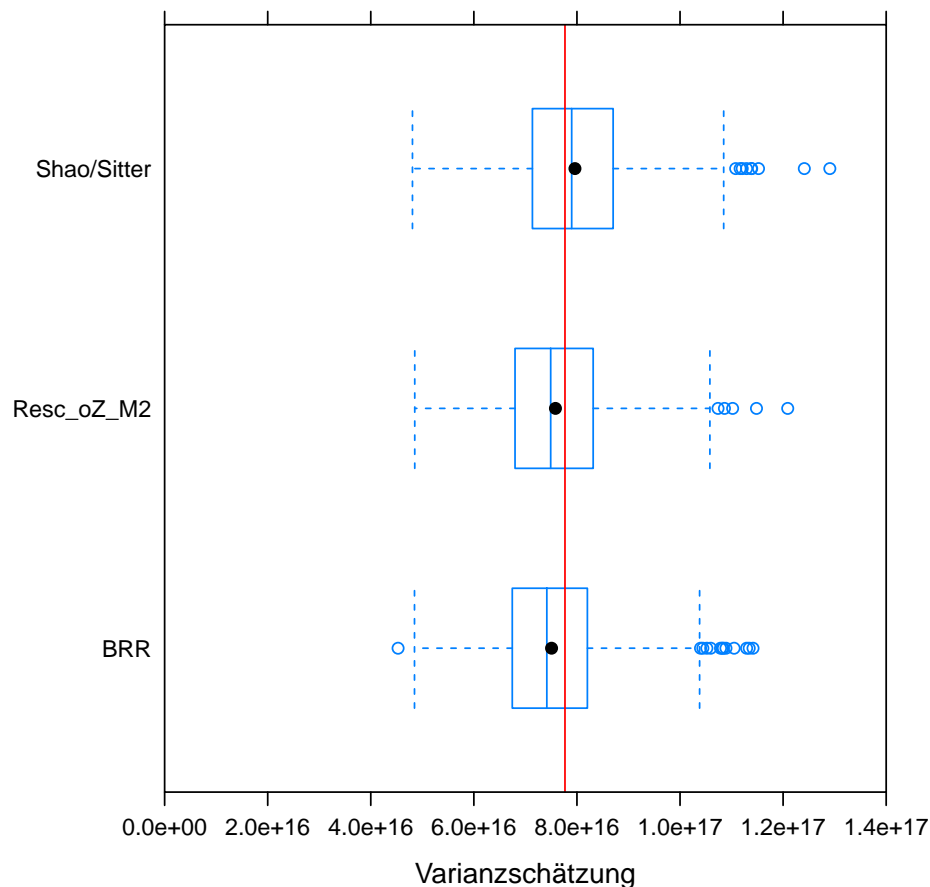


Abbildung 5.20: Erzeugung variabler fehlender Werte in der Grundgesamtheit in den gering besetzten Kategorien bei Mittelwertimputation, kleinen Auswahlätze und MAR

5.2.3.3 Vergleich unterschiedlicher Ausfallmechanismen

Abbildung 5.21 zeigt den relativen Bias der Schätzungen der drei Ausfallmechanismen MCAR, MAR und NMAR, wie in Abschnitt 2.1 behandelt.²⁴ Für alle drei Szenarien liegt die Entstehung variabler fehlender Werte in der Grundgesamtheit, Mittelwertimputation

²³Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.7 im Anhang A.

²⁴Vgl. hierzu die Parameterkonstellationen der Szenarien in der Datensatzbeschreibung in Abschnitt A.2.2.8 im Anhang A.

und kleine Auswahlätze vor. Die Ergebnisse für den Fall von MAR sind aus Abschnitt 5.2.1.1 bekannt. Die drei Resampling-Verfahren liefern in allen drei Fällen die bekannten Ergebnisse. Der BRR und der Rescaling-Bootstrap ohne Zurücklegen führen zu guten Ergebnissen, der Shao und Sitter-Bootstrap geht mit Überschätzungen der Varianz einher. Wie in den Abschnitten 3.6 und 5.2.1.1 beschrieben, ist der direkte Varianzschätzer bei Mittelwertimputation vom Ausfallmechanismus abhängig. Für MCAR ist er fast unverzerrt, für die beiden anderen Fälle sind deutliche Unterschätzungen zu sehen. Dieser Varianzschätzer berücksichtigt keine Hilfsinformationen und kann daher im Falle von MAR nicht verwendet werden.

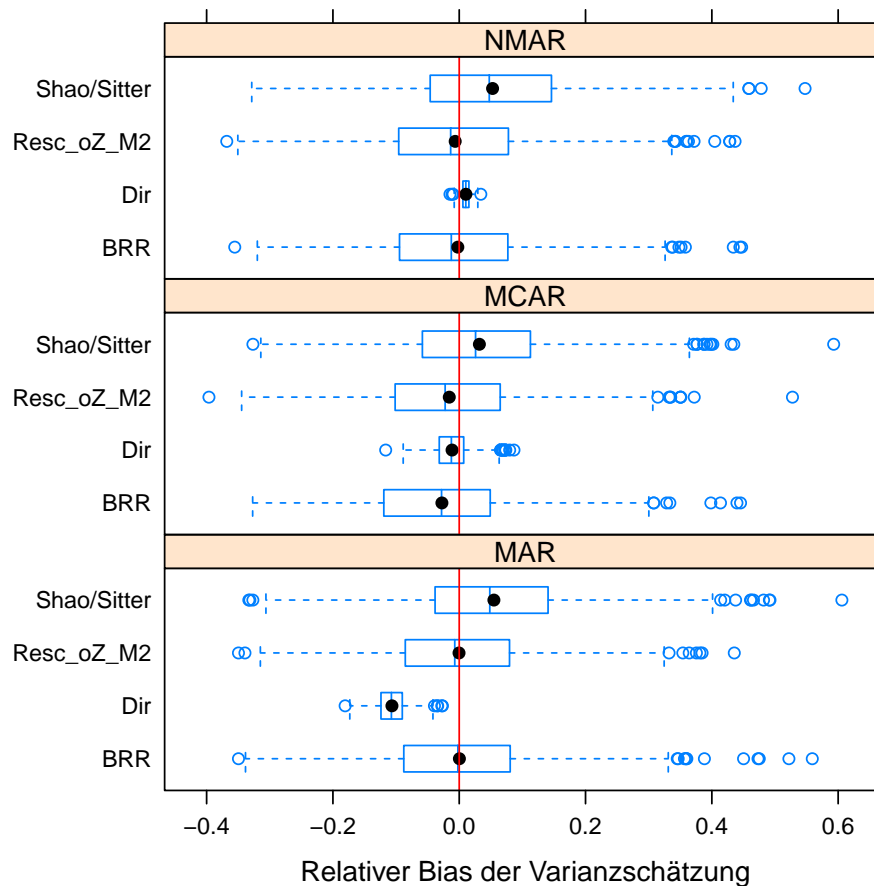


Abbildung 5.21: Vergleich der verschiedenen Ausfallmechanismen für Mittelwertimputation, die Entstehung variabler fehlender Werte in der Grundgesamtheit und kleine Auswahlätze

Im Falle von MCAR ist es jedoch auffällig, dass kleinere Unterschätzungen der Varianz auftreten, selbst für den Rescaling-Bootstrap und den BRR, obwohl die beiden anderen Antwortmechanismen deutlich komplexer sind. Dies hängt mit den Erläuterungen des vorherigen Abschnittes zusammen. Entscheidend ist wiederum die am geringsten besetzte Ausprägung der Hilfsvariable Haushaltstyp, damit die Kategorie 2. Im Falle von MAR wurden fehlende Werte so erzeugt, dass der Ausfall in anderen Kategorien auftritt (vgl. Abbildung 5.4). Auch im Falle von NMAR ist der Ausfall in dieser Kategorie nicht zu groß. Bei MCAR ist der Ausfall jedoch in allen Kategorien ungefähr gleich verteilt. Da die gesamte Anzahl an fehlenden Werten in allen Szenarien annähernd gleich groß ist,

ist der Ausfall in diesem Szenario in Kategorie 2 größer als für MAR und NMAR. Die Ergebnisse des vorherigen Abschnittes haben schon gezeigt, dass bei höherem Ausfall dieser Kategorie Ungenauigkeiten bei der Varianzschätzung einhergehen. Und so verhält es sich auch in diesem Szenario.

Die Varianzschätzungen im Falle von NMAR liefern mit Ausnahme des Shao und Sitter-Bootstrap unter den gegebenen Parameterkonstellationen der Simulationsstudie zwar gute Ergebnisse. Es ist jedoch zu beachten, dass der Punktschätzer hier verzerrt ist (vgl. Abschnitt A.3.1 im Anhang A) und daher spielen Varianzschätzungen nur noch eine untergeordnete Rolle.²⁵

5.2.3.4 Kombination von Imputationsverfahren

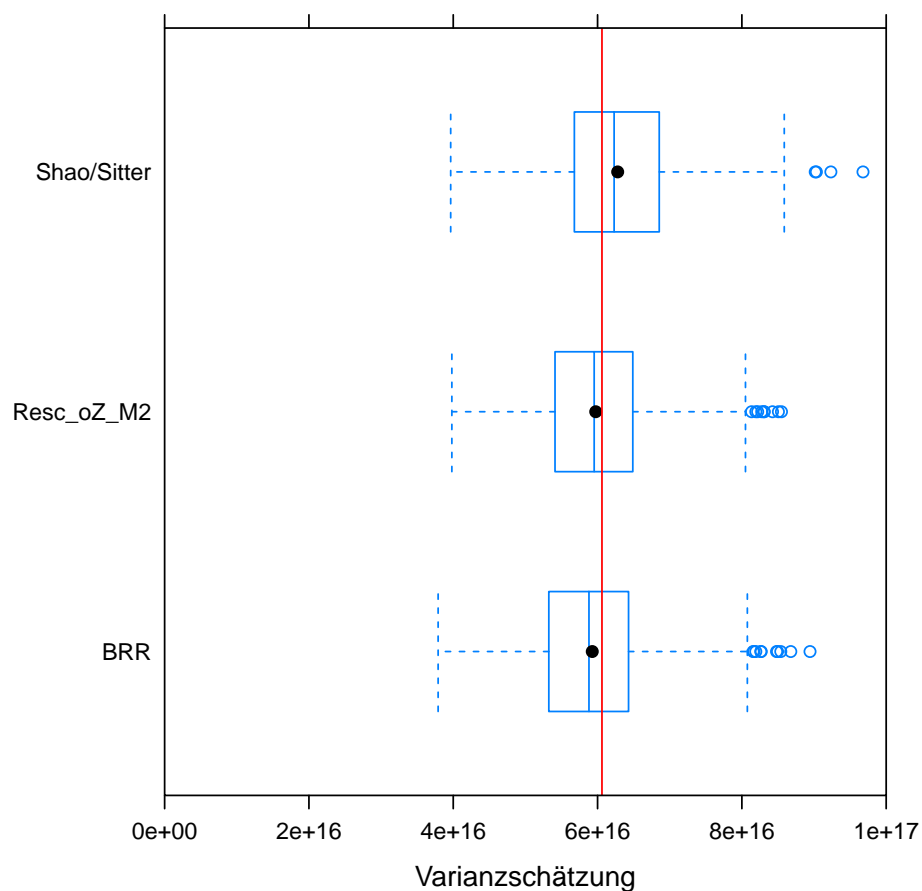


Abbildung 5.22: Kombination aus Mittelwert- und Hot-Deck-Random-Imputation bei kleinen Auswahlstößen und Entstehung variabler fehlender Werte in der Grundgesamtheit

In der Praxis kann es häufig notwendig sein, auf mehr als ein Imputationsverfahren zurückzugreifen, welche in den verschiedenen Imputationsklassen angewendet werden. In solchen Fällen, insbesondere in Kombination mit komplexen Stichprobendesigns, bieten sich große

²⁵Eindeutige Aussagen zur Qualität der Varianzschätzer können aufgrund der gegebenen Verzerrung des Punktschätzers auch nicht getroffen werden.

Vorteile durch die Anwendung von Resampling-Verfahren (vgl. Abschnitt 4.1.4). Abbildung 5.22 zeigt die Kombination von unterschiedlichen Imputationsverfahren.²⁶ In den Kategorien 1, 3 und 5 der Variablen Haushaltstyp erfolgt die Anwendung von Mittelwertimputation, in den Kategorien 2, 4 und 6 von Hot-Deck-Random-Imputation. Auch hier zeigen sich die bekannten Resultate. Der Rescaling-Bootstrap und der BRR führen zu fast unverzerrten Ergebnissen. Kleinere Unterschätzungen sind der Komplexität durch die Kombination der Imputationsverfahren geschuldet. Für die beiden Resampling-Methoden zeigen sich insgesamt sehr gute Schätzungen, was die Vorteilhaftigkeit der Anwendung solcher Verfahren im Falle der Kombination von Imputationsmethoden unterstreicht. Beim Shao- und Sitter-Bootstrap zeigen sich, wie erwartet, auch in diesem Szenario kleinere Überschätzungen.

5.2.3.5 Fazit zu den Spezialuntersuchungen

Der Abschnitt hat die Vorteilhaftigkeit der Anwendung von Resampling-Verfahren im Falle der Kombination von Imputationsverfahren gezeigt. Probleme treten hingegen auf, wenn fehlende Werte schwerpunktmäßig in den gering besetzten Kategorien erzeugt werden. Es wurde außerdem auf die Notwendigkeit der Berücksichtigung von Korrekturfaktoren zur Erfassung der Komponente bezüglich fehlender Werte hingewiesen. Dieser Sachverhalt wird im nächsten Abschnitt vertieft.

5.2.4 Große Auswahlsätze

Im Rahmen dieser Arbeit wurde ausführlich auf die Problematiken der Anwendung von Resampling-Verfahren bei großen Auswahlsätzen hingewiesen. Zum einen gilt es, wie in Abschnitt 4.7 beschrieben, bestimmte Korrekturfaktoren zu berücksichtigen, um die Varianzkomponente bezüglich fehlender Werte V_{NR}^1 bzw. \tilde{V}_{NR}^1 ²⁷ zu erfassen sowie bestimmte Adjustierungen im Falle zufälliger Imputationsverfahren vorzunehmen. Zum anderen gewinnen beim Rescaling-Bootstrap oder beim BRR nach (4.31) auch die nicht für die Stichprobe gezogenen Elemente bei der Berechnung des Punktschätzers der Resampling-Methode an Bedeutung. Da die Berechnung von Resampling-Methoden bei großen Auswahlsätzen mit größeren Laufzeiten einhergehen, werden in diesem Abschnitt nur noch der Rescaling-Bootstrap ohne Zurücklegen und der BRR betrachtet.

5.2.4.1 Mittelwertimputation

Abbildung 5.23 zeigt die Ergebnisse der Varianzschätzung für den Fall großer Auswahlsätze und der Verwendung von Mittelwertimputation.²⁸ Das erste Panel beinhaltet den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit, das zweite die Entstehung fehlender Werte in der Stichprobe und das dritte den eines festen Response-Vektors in der Grundgesamtheit. Beim Rescaling-Bootstrap wird die Unterscheidung getroffen, ob eine

²⁶In diesem Szenario gelten die Parameterkonstellationen der Mittelwertimputation aus Abschnitt A.2.2.2 im Anhang A.

²⁷Wie Abschnitt 4.7.3 beschrieben, beziehen sich die Korrekturfaktoren auf den Fall der Entstehung fehlender Werte in der Grundgesamtheit und die in diesem Fall definierte Varianzkomponente bezüglich fehlender Werte. Die Korrekturfaktoren werden in der Simulationsstudie auch für den Fall der Entstehung fehlender Werte in der Stichprobe verwendet. Dieser Sachverhalt wird im Rahmen der Varianzzerlegung in Abschnitt 5.3.1 näher erläutert.

²⁸Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.9 im Anhang A.

Reimputation der imputierten Werte innerhalb der gezogenen Elemente oder auch innerhalb der nicht gezogenen Elemente durchgeführt wird. Werden nur gezogene Einheiten berücksichtigt, wird weiterhin von Modifikation 1 gesprochen und es erfolgt eine Kennzeichnung mit M1. Wird eine Reimputation separat innerhalb der gezogenen und nicht gezogenen Elemente durchgeführt, stellt dies weiterhin die Modifikation 2 dar mit dem entsprechenden Kürzel M2. Außerdem erfolgt eine Differenzierung bezüglich der Verfahren, welche einen Korrekturfaktor zur Erfassung der Varianzkomponente bezüglich fehlender Werte V_{NR}^1 berücksichtigen und denjenigen, welche diese Größen nicht beinhalten. Für den Fall der Einbeziehung solcher Faktoren besitzt die Methode das entsprechende Kürzel KF.

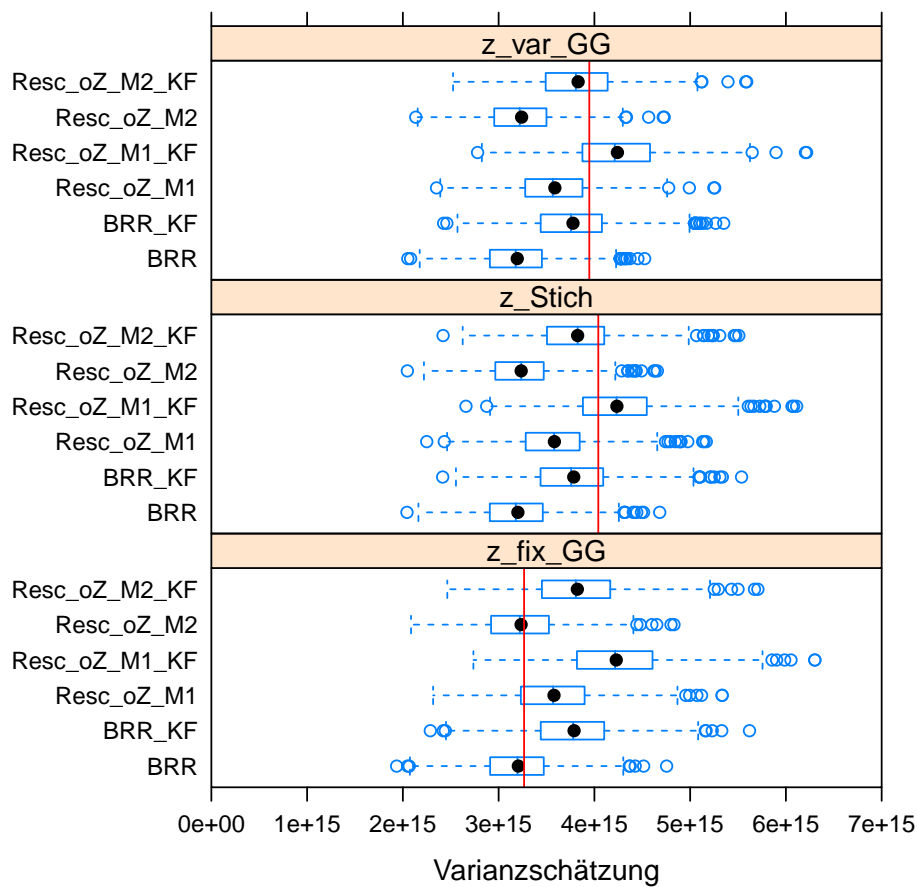


Abbildung 5.23: Vergleich der Varianzschätzer im Fall großer Auswahlsätze für Mittelwertimputation und MAR

Bei einem Vergleich der drei Fälle in Abbildung 5.23 kann festgestellt werden, dass sich die Ergebnisse der Fälle der Entstehung variabler fehlender Werte in der Grundgesamtheit (z_var_GG) und der Stichprobe (z_Stich) sehr ähnlich sind. Dies gilt sowohl für den Benchmark wie auch bezüglich der Varianzschätzer. Hier spielen getätigte Annahmen aus Abschnitt 3.2 ein Rolle. Dieser Sachverhalt wird ausführlich im Rahmen Varianzzerlegung in Abschnitt 5.3.1 erörtert. Bei Vorliegen eines fixen Response-Vektors in der Grundgesamtheit (z_fix_GG) ist die Varianz aufgrund des Wegfallens der Varianzkomponente bezüglich fehlender Werte V_{NR}^1 naturgemäß kleiner. Hier liefert die zweite Modifikation ohne Verwendung des Korrekturfaktors für den Rescaling-Bootstrap und den BRR die beste Lösung und unverzerrte Schätzergebnisse. Bezüglich der beiden Fälle der Entste-

hung von variablen fehlenden Werten muss der Korrekturfaktor berücksichtigt werden. Diese Resultate bekräftigen die Ausführungen der Abschnitte 3.4.2, 4.7.1 und 4.7.3 der Arbeit. Bei einem fixen Response-Vektor fällt die Varianzkomponente bezüglich fehlender Werte V_{NR}^1 nicht an. Wird der Korrekturfaktor eingerechnet, wird unterstellt, dass diese Varianzkomponente vorhanden ist, obwohl sie in diesem Fall überhaupt nicht auftritt. So kommt es zu einer Überschätzung der Varianz, wie im unteren Panel der Abbildung zu sehen ist. Bei den beiden Szenarien mit variablem Response-Vektor im ersten und zweiten Panel muss der Korrekturfaktor aus (4.50) verwendet werden. Die Varianzkomponente bezüglich fehlender Werte V_{NR}^1 besitzt durch den großen Gesamtauswahlsatz einen relativ hohen Anteil. Bei Nichtberücksichtigung des Korrekturfaktors wird die Varianz drastisch unterschätzt. Durch die Berücksichtigung des Korrekturfaktors kann diese Unterschätzung deutlich verringert werden. Es existieren hier aber immer noch kleinere Abweichungen vom Benchmark. Diese können zum einen mit dem Untersuchungsmerkmal erklärt werden, welche das Einkommen darstellt. Hier ist die Verteilung sehr schief wie in Abbildung 5.1 gezeigt wurde. Im Rahmen der mehrstufigen Designs wird eine normalverteilte Variable verwendet werden, wodurch bessere Ergebnisse resultieren. Diese Schiefe scheint sich auch eher auf die Schätzung der Varianzkomponente bezüglich fehlender Werte V_{NR}^1 auszuwirken, da im Falle eines festen Response-Vektors unverzerrte Schätzungen vorhanden sind. Zum anderen sind die Ergebnisse vergleichbar mit den Resultaten der Simulationsstudie von MASHREGHI et al. (2014). Die Größenordnung der relativen Verzerrung liegt ebenfalls zwischen minus 2% und 6%. Wie in Abschnitt 5.1.6 beschrieben, konnten zudem auch bestimmte Annahmen wie der uniforme Antwortmechanismus oder die einfache Zufallsstichprobe nicht eingehalten werden. Dies kann ebenfalls einen Anteil an der Verzerrung besitzen. Aber trotz dieser Problematiken kann der relative Bias im Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit beim Rescaling-Bootstrap durch die Berücksichtigung des Korrekturfaktors von fast 18% auf 3% reduziert werden.

Der zweite wichtige Punkt im Rahmen großer Auswahlsätze ist die Überprüfung der Notwendigkeit der zusätzlichen Reimputation der imputierten Werte der nicht für die Substichprobe gezogenen Einheiten, wie in dieser Arbeit vorgeschlagen. Diese Einheiten besitzen in diesem Fall auch einen Einfluss bei der Berechnung des Punktschätzers der einzelnen Replikationen. Die Schätzungen in Abbildung 5.23, insbesondere für den Fall eines fixen Response-Vektors im unteren Panel zeigen, dass eine Reimputation der imputierten Werte der nicht für die Substichprobe gezogenen Elemente unabdingbar ist. Nur für die Modifikation 2 des Rescaling-Bootstrap ohne Korrekturfaktor (*Resc_oZ_M2*) resultieren in diesem Szenario unverzerrte Schätzer. Wird die Reimputation der imputierten Werte der nicht für die Substichprobe gezogenen Elemente ausgelassen, wie für den Schätzer *Resc_oZ_M1*, entstehen große Überschätzungen der Varianz. Dieser Sachverhalt wurde in Abschnitt 4.5.4 näher erläutert. Durch die Vernachlässigung der Reimputation bestimmter Einheiten wird der ursprüngliche Imputationsprozess nicht korrekt wiedergegeben. Für fehlende Werte wird in Abhängigkeit, ob sie gezogen wurden, entweder der reimputierte Wert oder der Imputationswert der ursprünglichen Stichprobe verwendet. Über die Substichproben hinweg entsteht dadurch zu viel Variation, welche auf diese Art im ursprünglichen Imputationsprozess nicht enthalten war. Dadurch entstehen Verzerrungen. Daher muss bei großen Auswahlsätzen die Reimputation der Imputationswerte nicht für die Substichprobe gezogener Einheiten erfolgen.

Der Vergleich zwischen BRR und Rescaling-Bootstrap bei großen Auswahlsätzen lässt keine großen Unterschiede erkennen. Die Schätzer *Resc_oZ_M2* und *BRR* einerseits sowie *Resc_oZ_M2_KF* und *BRR_KF* andererseits weichen nur marginal voneinander ab. Die Diffe-

renzen können das Resultat von Zufälligkeiten der Monte-Carlo-Simulation sein. Aufgrund dieser geringen Unterschiede zwischen BRR und Rescaling-Bootstrap ohne Zurücklegen wurde auf die Angabe des BRR hinsichtlich der Modifikation 1 verzichtet. Bezüglich des Rescaling-Bootstrap wurde hier schon gezeigt, dass die Variante suboptimal ist.

5.2.4.2 Hot-Deck-Random-Imputation

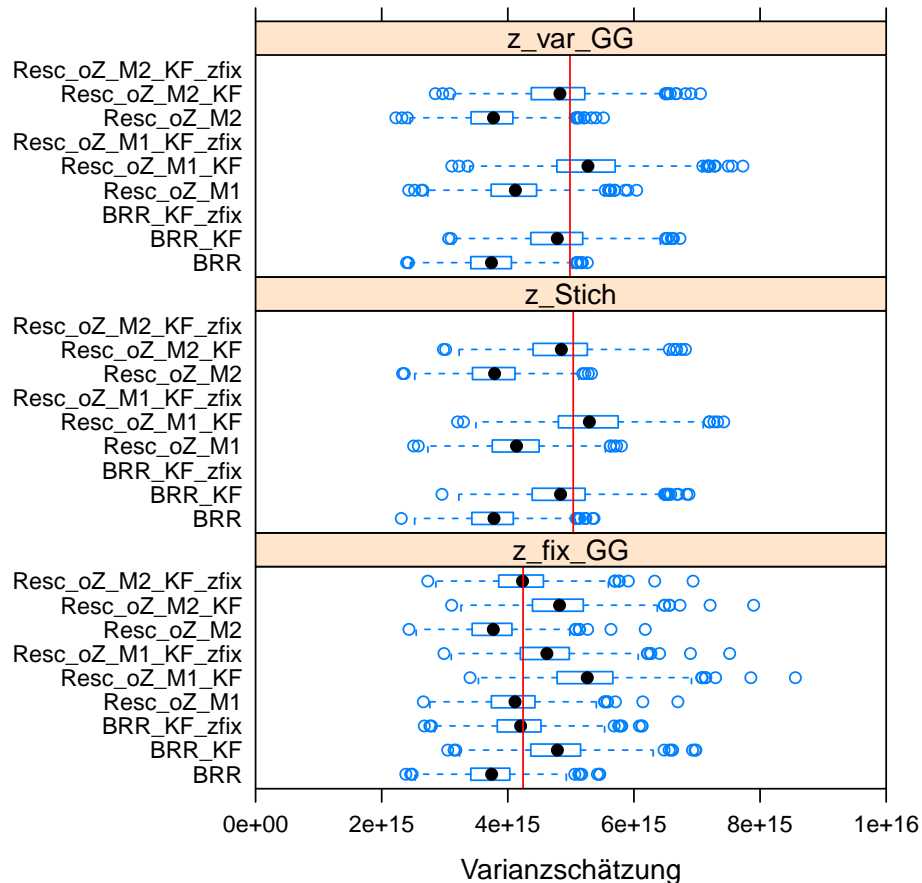


Abbildung 5.24: Vergleich der Varianzschätzer im Fall großer Auswahlsätze für Hot-Deck-Random-Imputation und MAR

Auch für die Hot-Deck-Imputation in Abbildung 5.24 können für die Fälle variabler fehlender Werte die zuvor getätigten Aussagen übernommen werden.²⁹ Beim ersten und zweiten Panel ist zu sehen, dass durch die Verwendung des Korrekturfaktors aus (4.59) die Verzerrung deutlich reduziert wird. Es entsteht die Varianzkomponente bezüglich fehlender Werte \tilde{V}_{NR}^1 und mit Hilfe des Korrekturfaktors kann sie erfasst werden. Der Fall eines fixen Response-Vektors in der Grundgesamtheit, dargestellt im dritten Panel, hat im Zusammenhang mit zufälligen Imputationsverfahren eine Besonderheit. Wie im deterministischen Fall fällt die Komponente \tilde{V}_{NR}^1 nicht an. An den Schätzern *Resc_oZ_M2_KF* und *BRR_KF* ist zu sehen, dass bei Berücksichtigung des Korrekturfaktors die Varianz überschätzt wird. Wird dieser jedoch ausgelassen, wie bei den Schätzern *Resc_oZ_M2* und

²⁹In diesem Szenario gelten die Parameterkonstellationen der Mittelwertimputation aus Abschnitt A.2.2.9 im Anhang A.

BRR wird die Varianz unterschätzt. Dies liegt an der schlechten Approximation der gemeinsamen Verteilung aus Stichprobe und Donoren, wie in 4.7 beschrieben wurde. Im Rahmen dieser Arbeit wurde für den Fall eines festen Response-Vektors in der Grundgesamtheit und Anwendung von Hot-Deck-Random-Imputation in (4.65) ein modifizierter Korrekturfaktor angegeben. Dieser berücksichtigt, dass die Komponente \tilde{V}_{NR}^1 nicht auftritt, jedoch die schlechte Approximationsgüte kompensiert werden muss. Wird dieser Faktor beim Rescaling-Bootstrap und beim BRR verwendet, gekennzeichnet durch *Resc_oZ_M2_KF_zfix* und *BRR_KF_zfix*, ist zu sehen, dass unverzerrte Schätzungen resultieren. Daher darf in diesem Fall nur mit dem angepassten Korrekturfaktor gearbeitet werden. Schätzer *Resc_oZ_M1_KF_zfix* bei welchem nur die für die Substichprobe gezogenen Imputationswerte adjustiert werden, führt erwartungsgemäß zu einer Überschätzung der Varianz. Ein Vergleich mit den Schätzern *Resc_oZ_M2_KF_zfix* und *BRR_KF_zfix* zeigt, dass auch hier eine Adjustierung der Imputationswerte der nicht für die Substichprobe gezogenen Einheiten erfolgen muss.

5.2.4.3 Fazit für große Auswahlsätze

Die Theorien bezüglich der Anwendung von Resampling-Verfahren bei großen Auswahlsätzen, wie insbesondere in den Abschnitten 4.5.4, 4.6.3 und 4.7.3 erläutert, werden durch die Simulationsstudie bestätigt. Es muss eine Reimputation bzw. Adjustierung der Imputationswerte auch der nicht für die Substichprobe gezogenen Elemente vollzogen werden. Ansonsten kommt es zu deutlichen Überschätzungen der Varianz. Außerdem gilt es, die Verwendung von Korrekturfaktoren zu beachten. Diese werden basierend auf den Simulationsergebnissen noch einmal in Tabelle 5.4 zusammengefasst. Es werden die Imputationsmethoden Mittelwertimputation (*MW*) und Hot-Deck-Random-Imputation (*HDRI*) unterschieden. Außerdem werden die verschiedenen Fälle der Entstehung fehlender Werte betrachtet, sowie eine Unterscheidung vorgenommen, ob kleine oder große Auswahlsätze (*Ausw. klein* bzw. *Ausw. groß*) vorliegen.

Imp-Meth	Ausw. groß, Fall 1 u. 2	Ausw. groß, Fall 3	Ausw. klein
MW	KF nach (4.50)	Kein KF	Kein KF
HDRI	KF nach (4.59)	KF nach (4.65)	Kein KF

Tabelle 5.4: Notwendigkeit der Berücksichtigung von Korrekturfaktoren (KF)

5.3 Varianzzerlegung bezüglich der einfachen Zufallsstichprobe

Ziel dieses Abschnittes ist die Zerlegung der gesamten Varianz eines Schätzers bei Auftreten von fehlenden Werten und deren anschließender Imputation in ihre verschiedenen Komponenten. Diese Bestandteile der Varianz beziehen sich auf die Ausführungen in Abschnitt 3.4.2. Durch die Zerlegung der Varianz in einzelne Komponenten sollen auch die Auswirkungen der Variation bestimmter Parameter wie des Auswahlsatzes, der Imputationsmethode, die unterschiedlichen Fälle der Entstehung fehlender Werte oder der Ausfallrate auf diese Bestandteile sichtbar gemacht werden. Das Ziel ist hier die Überprüfung der Aussagen aus Abschnitt 3.4.3. Wichtig ist, dass die Variation eines Parameters unter größtmöglicher Konstanz der übrigen Parameter vorgenommen wird, um den Einfluss

zweifelsfrei zuordnen zu können. Im Basisszenario wird von einem mittleren Auswahlsatz, mittlerer Ausfallrate und der Erzeugung fehlender Werte in der Grundgesamtheit ausgegangen (vgl. hierzu Abschnitt 5.1.5).

Die einzelnen Komponenten wurden über eine Monte-Carlo-Simulation berechnet, wie in Abschnitt 5.1.3 erläutert. Kleinere Ungenauigkeiten sind aufgrund eines Monte-Carlo-Fehlers möglich. Im Rahmen dieses Abschnittes wird auf die einfache Zufallsstichprobe, MAR sowie die in 5.1.3 beschriebene Grundgesamtheit abgestellt.³⁰

5.3.1 Varianzzerlegung bezüglich der drei Fälle an fehlenden Werten und Mittelwertimputation

Wird zunächst eine Varianzzerlegung hinsichtlich der drei in dieser Arbeit bekannten Fälle der Entstehung fehlender Werte bei Mittelwertimputation vorgenommen, erfolgt eine Unterscheidung in Erzeugung variabler fehlender Werte in der Grundgesamtheit (z_var_GG), Erzeugung fehlender Werte in der Stichprobe (z_Stich) und eines fixen Response-Vektors in der Grundgesamtheit (z_fix_GG). Dies geschieht in Abbildung 5.25.

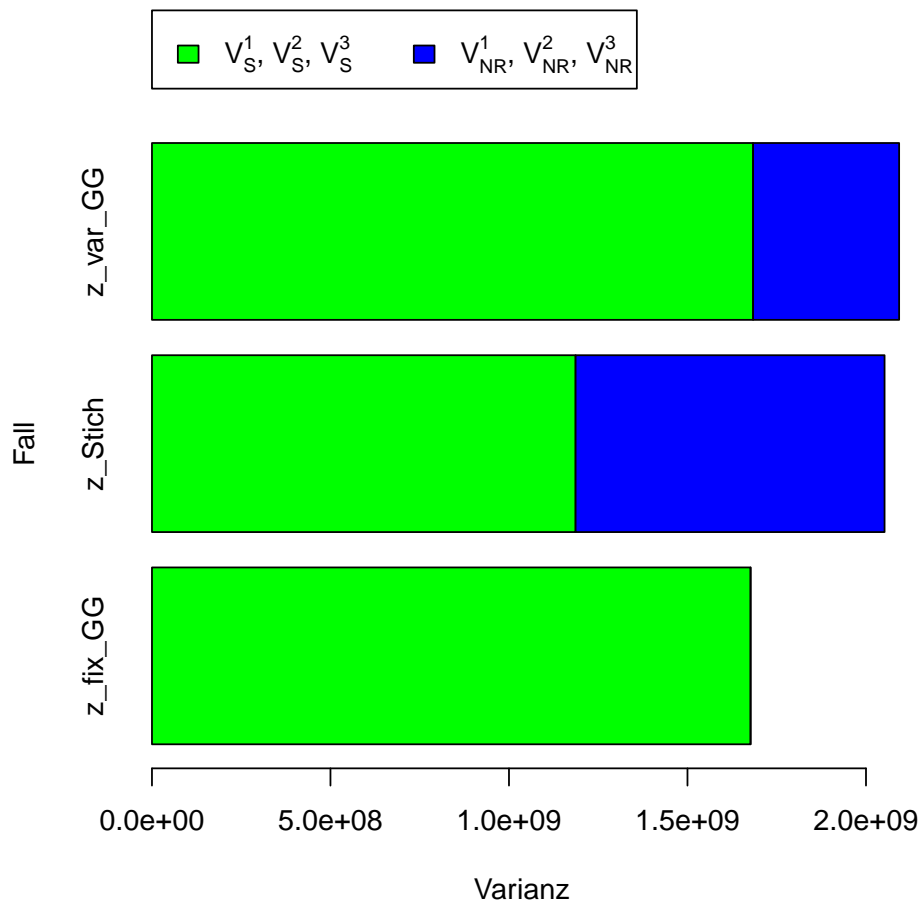


Abbildung 5.25: Varianzzerlegung hinsichtlich der drei Fälle an fehlenden Werten für Mittelwertimputation, mittlerem Auswahlsatz und mittlerer Rate an fehlenden Werten

³⁰Die Parameterkonstellationen im Falle der Varianzzerlegung für die einfache Zufallsstichprobe werden in Abschnitt A.2.2.10 im Anhang A gegeben.

Der große Unterschied zwischen der Entstehung variabler und fixer fehlender Werte in der Grundgesamtheit ergibt sich dadurch, dass bei letzterem die Varianzkomponente bezüglich fehlender Werte V_{NR}^1 nicht auftritt. Die Varianz besteht nur aus der Stichprobenkomponente V_S^3 und daher tritt nur der grüne Balken in Erscheinung. Wird der Response-Vektor von Simulations- zu Simulationsdurchlauf in z_var_GG neu generiert, kommt die Komponente V_{NR}^1 und damit auch der blaue Balken hinzu. Diese Komponente bildet hier jedoch im Vergleich zur Stichprobenkomponente V_S^1 nur einen kleinen Anteil, selbst bei einem Auswahlsatz von 50 %. Zwischen der Entstehung variabler fehlender Werte in der Grundgesamtheit und in der Stichprobe sind in der absoluten Höhe nur geringfügige Unterschiede zu sehen, welche das Resultat von Zufälligkeiten der Monte-Carlo-Simulation sein können. In Abschnitt 3.4.2 wurde beschrieben, dass sich diese beiden Fälle hinsichtlich der Zerlegung der gesamten Varianz unterscheiden. Dies betrifft die inneren wie äußeren Vorgänge sowie gemäß welcher Größen die Konditionierung der inneren Vorgänge erfolgt. In Abbildung 5.25 ist dieser, bereits in den Formeln (3.23) und (3.24) dargestellte Sachverhalt visualisiert. Die beiden Fälle unterscheiden sich bezüglich des Anteils beider Komponenten deutlich. Die Stichprobenkomponente V_S^1 für den Fall der Entstehung fehlender Werte in der Grundgesamtheit ist größer als bei der Generierung fehlender Werte in der Stichprobe V_S^2 . Ein gegenteiliges Verhalten ist bei der Komponente bezüglich der fehlenden Werte V_{NR}^1 bzw. V_{NR}^2 ersichtlich.

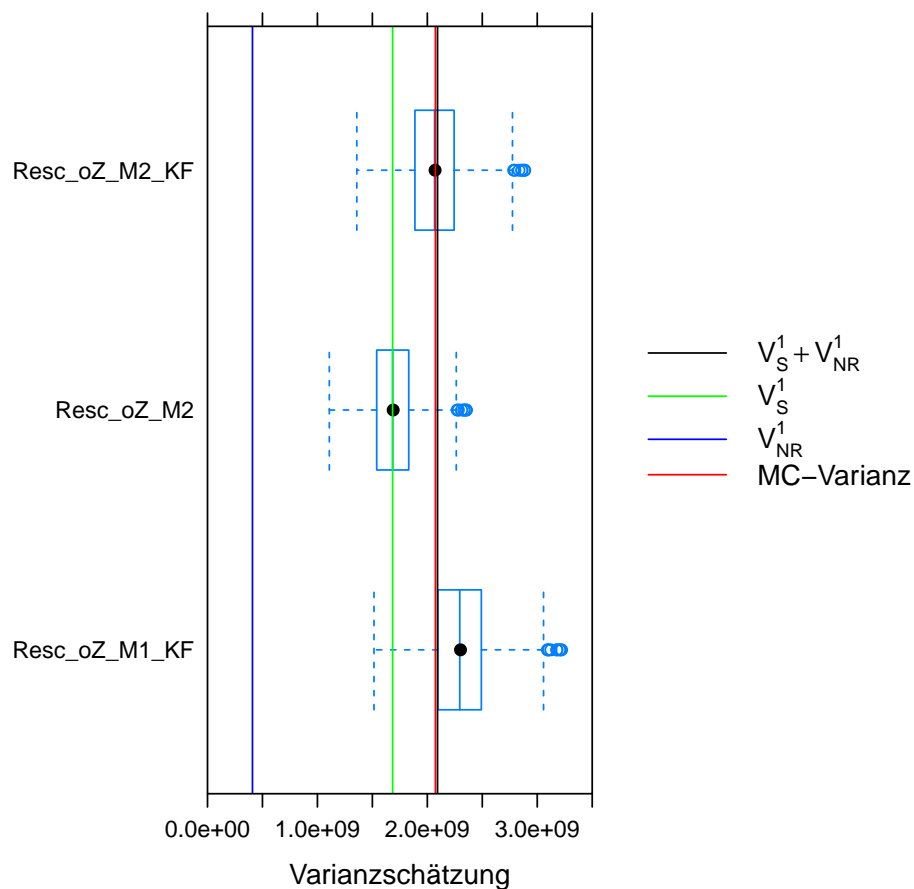


Abbildung 5.26: Abdeckung der einzelnen Varianzkomponenten durch den Rescaling-Bootstrap für einen mittleren Auswahlsatz, Mittelwertimputation und der Entstehung variabler fehlender Werte in der Grundgesamtheit

Diese berechneten Varianzkomponenten können verwendet werden, um noch einmal zu verdeutlichen welche Bestandteile der Varianz Resampling-Verfahren genau abdecken. Dies wird am Beispiel des Rescaling-Bootstrap ohne Zurücklegen gezeigt. Abbildung 5.26 zeigt den Fall variabler fehlender Werte in der Grundgesamtheit. Die einzelnen Versionen des Rescaling-Bootstrap unterscheiden sich je nachdem, ob ein Korrekturfaktor verwendet wird und ob nicht gezogene Werte auch reimputiert werden. Die Varianzkomponenten aus Abbildung 5.25 werden durch die verschiedenen Linien präsentiert. Die grüne Linie stellt die Komponente V_S^1 dar. Es ist zu sehen, dass der Rescaling-Bootstrap ohne Korrekturfaktor und mit Reimputation der gezogenen und nicht gezogenen Elemente (*Resc.oZ.M2*) gerade diese Komponente abdeckt. Die Varianzkomponente V_{NR}^1 wird hingegen nicht berücksichtigt. Um diese Komponente zu erfassen muss der Korrekturfaktor aus (4.50) berücksichtigt werden. Nur eine einzige Version des Rescaling-Bootstrap ist in der Lage, die Gesamtvarianz korrekt abzudecken. Wie zuvor ist es diejenige Methode, welche sowohl den Korrekturfaktor wie auch die separate Reimputation gezogener und nicht gezogener Elemente beinhaltet (*Resc.oZ.M2.KF*). Es existiert nur eine geringfügige Abweichung zwischen der Monte-Carlo-Varianz und der Summe der beiden Varianzkomponenten. Diese kann durch Zufälligkeiten der Monte Carlo Simulation entstehen. Wie zuvor überschätzt der Rescaling-Bootstrap mit ausschließlicher Reimputation der gezogenen Elemente (*Resc.oZ.M1.KF*). Hier entsteht, wie ausführlich erläutert, zu viel Variation über die Substichproben hinweg, durch die Verwendung entweder des imputierten oder des reimputierten Wertes in Abhängigkeit der Substichprobenziehung.

Wie bereits aus Abbildung 5.25 ersichtlich wurde, ergibt sich eine andere Zusammensetzung der Varianz, wenn die fehlenden Werte in der Stichprobe erzeugt werden. Dabei resultiert eine dem Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit vergleichbare Gesamtvarianz. Dieser Sachverhalt ist auch in Abbildung 5.27 zu sehen. Die Stichprobenkomponente wird kleiner, die Varianzkomponente bezüglich fehlender Werte größer. Die Auswirkungen auf die Resampling-Verfahren können an dieser Abbildung noch einmal detailliert beschrieben werden. Insbesondere zeigt sich, warum die Verwendung der Korrekturfaktoren zu einer unverzerrten Schätzung der Gesamtvarianz führt, obwohl diese für den Fall variabler fehlender Werte in der Grundgesamtheit hergeleitet worden sind. Der Vergleich zu Abbildung 5.26 zeigt, dass der Rescaling-Bootstrap unabhängig von dem Ort der Entstehung fehlender Wert in gleicher Höhe anfällt. Er deckt genau die gleiche Quantität ab, obwohl sich die Varianzkomponenten unterscheiden. Ohne Korrekturfaktor schätzt er im Fall variabler Erzeugung fehlender Werte in der Grundgesamtheit in Abbildung 5.26 die Stichprobenkomponente V_S^1 unverzerrt. Fallen fehlende Werte jedoch in der Stichprobe an, kann eine solche genaue Abgrenzung nicht mehr vollzogen werden, da der Anteil der Stichprobenkomponente V_S^2 geringer wird, sich an den Schätzungen des Rescaling-Bootstrap aber im Vergleich zum vorherigen Fall nicht viel ändert. Die Methode deckt daher nicht mehr die Komponente V_S^2 ab. Die Gesamtvarianz ist in beiden Fällen ungefähr gleich hoch mit Ausnahme von Zufälligkeiten der Monte-Carlo-Simulation. Da der Rescaling-Bootstrap in beiden Fällen zu gleichen Schätzergebnissen führt und die Gesamtvarianz gleich groß ist, führt die Einbeziehung des Korrekturfaktors in die Varianzschätzmethode auch im Falle der Entstehung fehlender Werte in der Stichprobe zur Abdeckung der Gesamtvarianz.

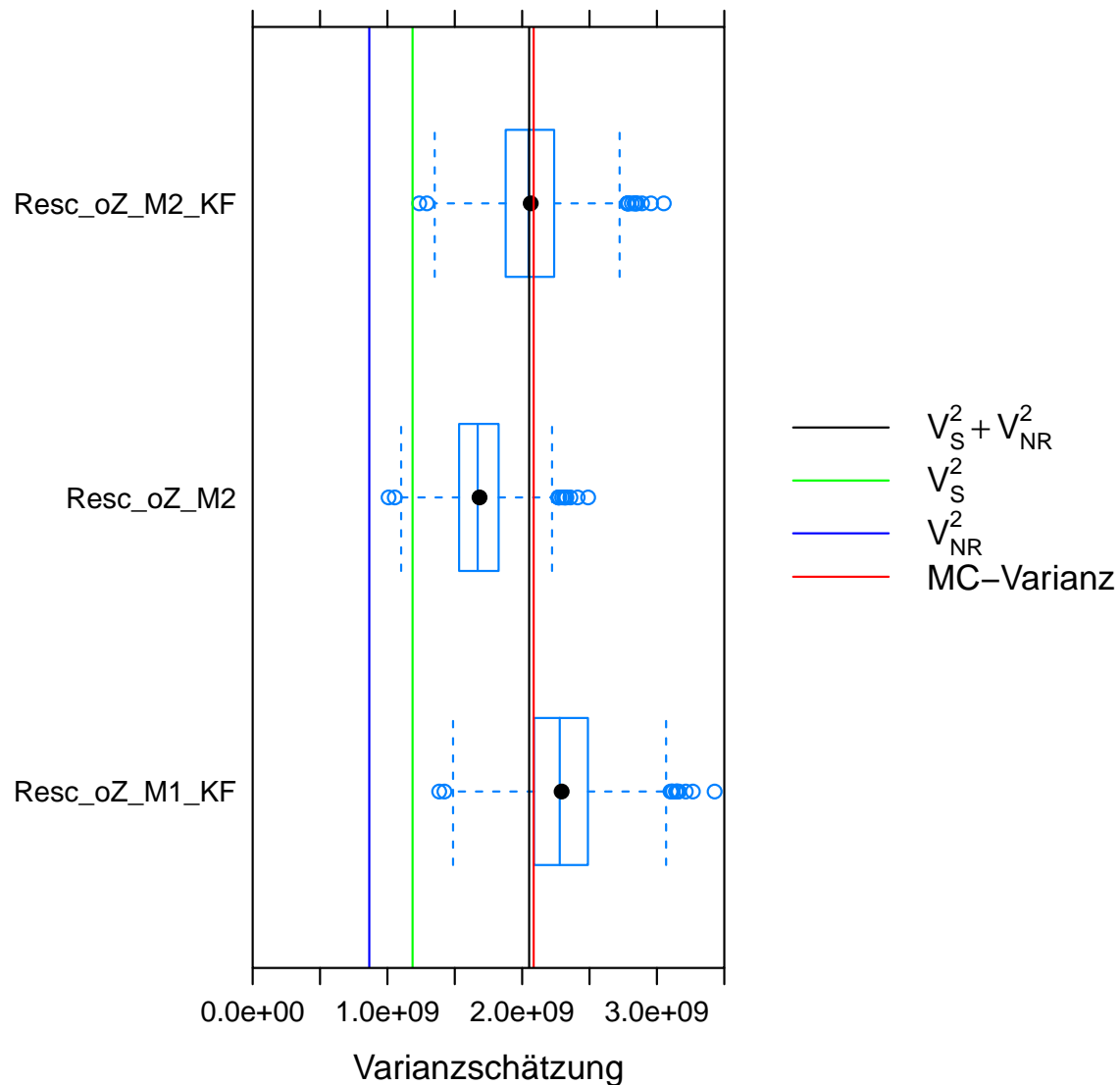


Abbildung 5.27: Abdeckung der einzelnen Varianzkomponenten durch den Rescaling-Bootstrap für einen mittleren Auswahlsatz, Mittelwertimputation und der Entstehung fehlender Werte in der Stichprobe

Es bleibt noch zu klären, warum der Rescaling-Bootstrap in beiden Fällen zu gleichen Schätzergebnissen führt. Grund hierfür ist zum einen, dass in beiden Fällen sowohl das Stichprobendesign wie auch der Ausfallmechanismus grundsätzlich gleich sind. Es existieren keine Abhängigkeiten zwischen der Stichprobenziehung und Erzeugung des Response-Vektors. Diese Annahme wurde in Abschnitt 3.2 getätigt und gilt auch für diese Simulationsstudie. Es unterscheidet sich lediglich der Ort der Entstehung fehlender Werte. Dadurch ergeben sich Stichproben- und Response-Vektoren, welche sehr ähnlich sind. Zudem wird der Rescaling-Bootstrap in beiden Fällen nach dem gleichen Vorgehen auf diese Vektoren angewendet. Unterschiede der Schätzungen durch die Resampling-Methode können damit nicht entstehen.

Im dritten Fall einmalig generierter fehlender Werte in der Grundgesamtheit in Abbildung 5.28 lassen sich die Erkenntnisse aus Abschnitt 3.4.2 im Detail nachvollziehen. Hier tritt die Varianzkomponente durch fehlende Werte V_{NR}^1 nicht auf. Die Gesamtvarianz entspricht

V_S^3 . Der Rescaling-Bootstrap ohne Berücksichtigung des Korrekturfaktors führt damit zu unverzerrten Ergebnissen. Durch Berücksichtigung des Korrekturfaktors, wird unterstellt, dass die Komponente V_{NR}^1 auftritt. Da dies aber nicht zutrifft, resultieren Überschätzungen bei der Berücksichtigung dieses Terms.

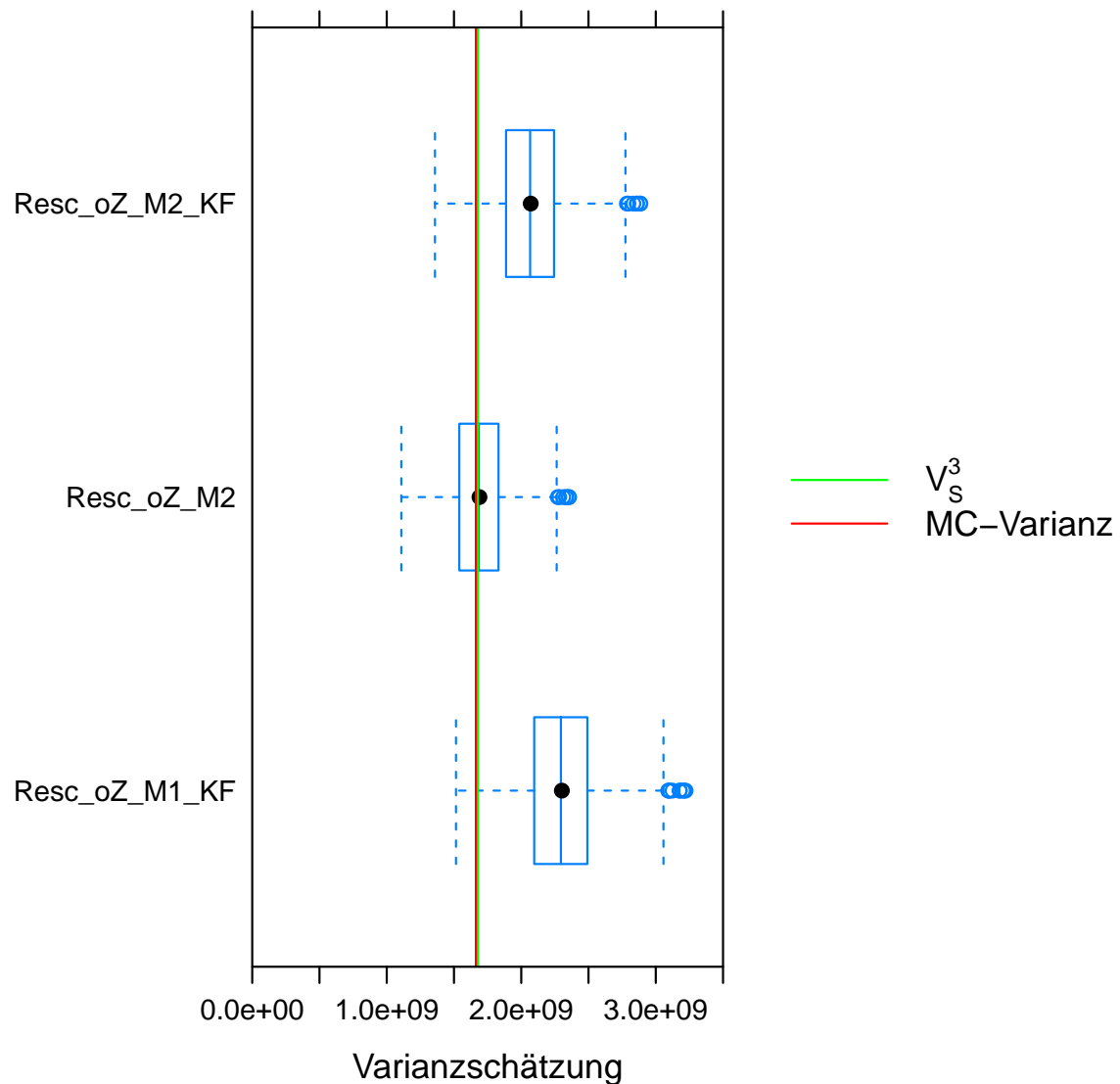


Abbildung 5.28: Abdeckung der einzelnen Varianzkomponenten durch den Rescaling-Bootstrap für einen mittleren Auswahlsatz, Mittelwertimputation und Erzeugung eines fixen Response-Vektors in der Grundgesamtheit

5.3.2 Varianzzerlegung bezüglich unterschiedlicher Auswahlätze bei Mittelwertimputation

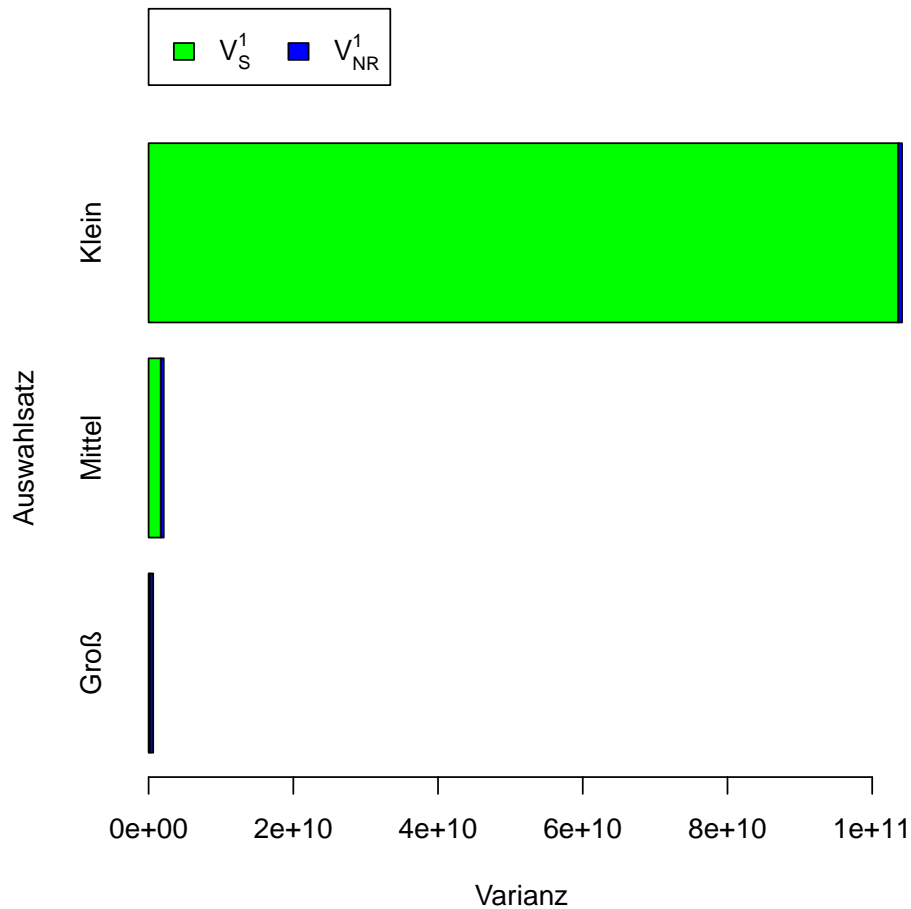


Abbildung 5.29: Varianzzerlegung hinsichtlich verschiedener Auswahlätze für Mittelwertimputation, Entstehung variabler fehlender Werte in der Grundgesamtheit und mittlerer Rate an fehlenden Werten

Die Variation des Auswahlssatzes in Abbildung 5.29 zeigt, dass der Anteil der Varianzkomponente bezüglich fehlender Werte V_{NR}^1 an der Gesamtvarianz von diesem Parameter abhängig ist. Die absolute Höhe dieser Komponente ist in allen drei Szenarien unabhängig vom Auswahlssatz ungefähr gleich groß. Dies wurde schon in Abschnitt 3.4.3 beschrieben. Die Höhe der Komponente V_{NR}^1 im Falle von Mittelwertimputation ist insbesondere vom Verhältnis fehlender Werte zu den beobachteten Werten abhängig, welches in allen drei Szenarien ungefähr gleich groß ist. Daher ändert sich die absolute Höhe der Komponente V_{NR}^1 kaum. Unterschiede können insbesondere aus Zufälligkeiten der Monte-Carlo-Simulation resultieren. Im Gegensatz hierzu hängt die Stichprobenkomponente V_S^1 sehr stark vom Auswahlssatz ab. Bei kleinem Auswahlssatz ist sie sehr groß und dominiert die Varianz. Bei großem Auswahlssatz ist sie sehr klein. Die Komponente V_{NR}^1 bekommt daher mit steigendem Auswahlssatz einen immer größer werdenden Anteil, da V_S^1 abnimmt, sie aber in der Höhe gleich bleibt. Es ist jedoch auch zu sehen, dass die Komponente V_{NR}^1 bei einem kleinen Auswahlssatz vernachlässigt werden kann.

5.3.3 Varianzzerlegung hinsichtlich verschiedener Ausfallraten bei Mittelwertimputation

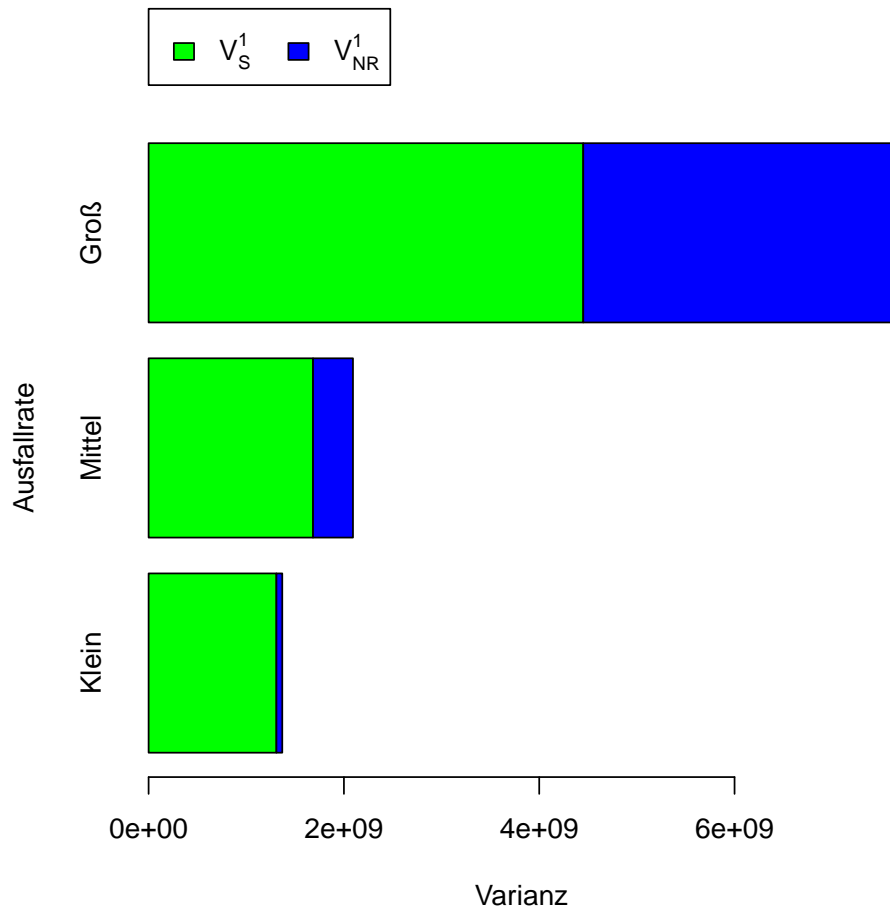


Abbildung 5.30: Varianzzerlegung hinsichtlich verschiedener Ausfallraten für Mittelwertimputation, Entstehung variabler fehlender Werte in der Grundgesamtheit und mittlerem Auswahlsatz

Bei diesem Vergleich findet eine Variation der Ausfallrate statt. Die übrigen Parameter sind gleich, es wird von der Entstehung variabler fehlender Werte in der Grundgesamtheit, Mittelwertimputation und einem Auswahlsatz von 50 % ausgegangen. In Abschnitt 3.4.3 wurden die Auswirkungen der Veränderung der Ausfallrate auf die Varianz des Schätzers näher erläutert. Wie dort beschrieben wurde, nimmt die Varianzkomponente bezüglich fehlender Werte V_{NR}^1 mit steigender Ausfallrate zu. Dies ist auch in der Abbildung 5.30 zu sehen. Aber nicht nur diese Komponente steigt an, sondern auch die Stichprobenkomponente V_S^1 . Dies liegt daran, dass diese Komponente die Stichprobenvarianz, konditioniert auf den Response-Vektor, enthält. Diese Komponente wird ebenfalls von den fehlenden Werten und den dadurch notwendigen Imputationen beeinflusst. Ceteris paribus nimmt mit größer werdender Anzahl auch hier die Unsicherheit zu und die Varianzkomponente wird größer. Mit steigender Rate an fehlenden Werten nehmen beide Komponenten zu und damit wird auch die Gesamtvarianz größer.

5.3.4 Varianzzerlegung hinsichtlich verschiedener Imputationsmethoden

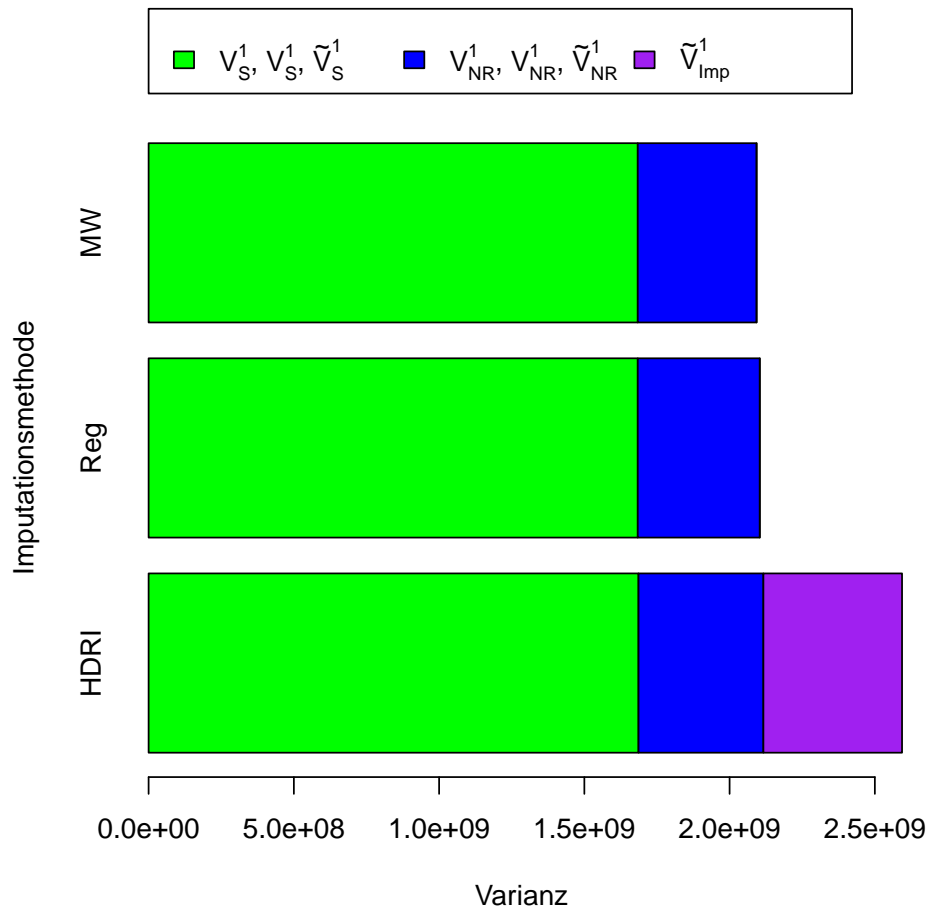


Abbildung 5.31: Varianzzerlegung hinsichtlich verschiedener Imputationsmethoden bei Entstehung variabler fehlender Werte in der Grundgesamtheit, mittlerer Ausfallrate und mittlerem Auswahlsatz

Abbildung 5.31 zeigt einen Vergleich zwischen den verschiedenen Imputationsmethoden Mittelwertimputation (*MW*), Regressionsimputation (*Reg*)³¹ und Hot-Deck-Random-Imputation (*HDRI*). Es zeigen sich keine größeren Unterschiede bezüglich der Varianzkomponenten, welche aufgrund von Variationen durch Stichproben oder fehlenden Werten resultieren V_S^1 bzw. \tilde{V}_S^1 und V_{NR}^1 respektive \tilde{V}_{NR}^1 . Wie zuvor dominiert bei der gegebenen Parameterkonstellation die Stichprobenkomponente V_S^1 bzw. \tilde{V}_S^1 . Durch die Ähnlichkeit ihrer Varianzkomponenten liegen zwischen Mittelwertimputation und der Regressionsimputation zudem kaum Unterschiede bezüglich der Gesamtvarianz vor. Dies ist hier der Fall, da bei beiden Methoden die gleichen Hilfsinformationen eingehen. Es wird lediglich die für die Simulation erzeugte kategoriale Variable verwendet. Im Rahmen der Mittelwertimputation geschieht dies über die Imputationsklassenbildung, bei der Regressionsimputation geht diese Hilfsvariable über das Regressionsmodell ein. Andere Hilfsvariablen

³¹Diese basiert hier im Gegensatz zu den vorherigen und nachfolgenden Abschnitten auf der kategorialen Hilfsvariable. Der Grund liegt in einer besseren Vergleichbarkeit mit der Varianzzerlegung der beiden anderen Imputationsmethoden.

werden nicht verwendet. Dadurch sind sich die Varianzen sehr ähnlich. Des Weiteren ist der Zusammenhang des Untersuchungsmerkmals zu der kategorialen Hilfsvariablen nicht so groß, dass durch die Regressionsimputation viel Variabilität unter den Imputationswerten entsteht, wie in Abschnitt 2.2.9 erläutert. Auch bei der Hot-Deck-Random-Imputation fallen die Komponenten \tilde{V}_S^1 und \tilde{V}_{NR}^1 in vergleichbarer Höhe zu den beiden deterministischen Imputationsverfahren an. Zur Imputationsklassenbildung wird auch hier nur die kategoriale Hilfsvariable verwendet. Wie erwartet nimmt jedoch die Gesamtvarianz zu. Es tritt zusätzlich die Imputationskomponente \tilde{V}_{Imp}^1 in Gestalt des lila Balkens auf. Diese ist das Resultat der zufälligen Auswahl der Imputationswerte. Hierdurch entsteht die zusätzliche Varianzkomponente.

5.3.5 Varianzzerlegung bezüglich der drei Fälle an fehlenden Werten für Hot-Deck-Random-Imputation

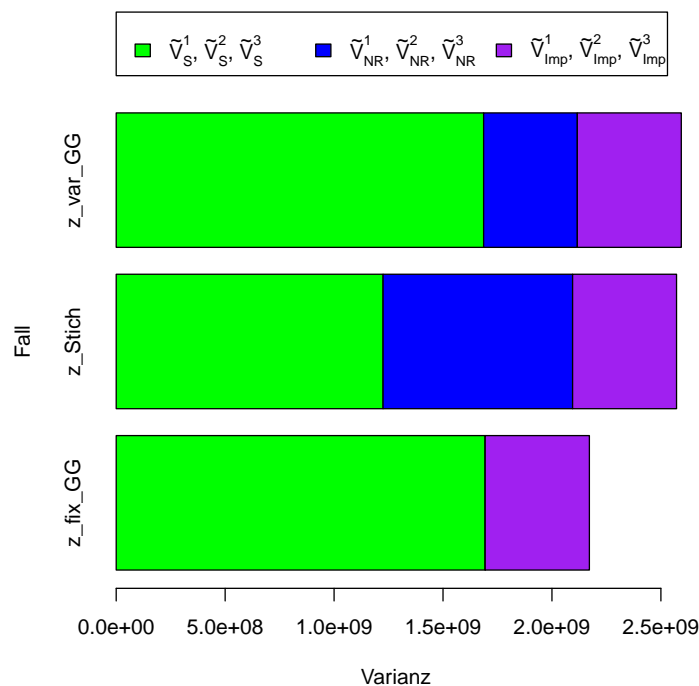


Abbildung 5.32: Varianzzerlegung hinsichtlich der drei Fälle an fehlenden Werten für Hot-Deck-Random-Imputation, mittlerer Ausfallrate und mittlerem Auswahlsatz

Abbildung 5.32 zeigt die Varianzzerlegung für die drei Fälle fehlender Werte für Hot-Deck-Random-Imputation. Die zentrale Erkenntnis ist, dass die Imputationskomponente in allen Fällen \tilde{V}_{Imp}^1 , \tilde{V}_{Imp}^2 und \tilde{V}_{Imp}^3 ungefähr gleich groß ist. Unter den gegebenen Bedingungen werden unabhängig vom Ort sehr ähnliche Stichproben- und Response-Vektoren erzeugt. Da die Variabilität durch die Imputation, als zentraler Bestandteil der Imputationskomponente, in allen drei Fällen gegeben beider Vektoren betrachtet wird, unterscheidet sie sich nicht stark. Die weiteren Resultate sind dem deterministischen Fall der Mittelwertimputation sehr ähnlich. Der Fall der Generierung variabler fehlender Werte in der Grundgesamtheit unterscheidet sich vom Fall eines festen Response-Vektors nur durch die zusätzliche Varianzkomponente bezüglich fehlender Werte \tilde{V}_{NR}^1 . Bei variablen

fehlenden Werten in der Grundgesamtheit ist gegenüber der Erzeugung fehlender Werte in der Stichprobe die Komponente \tilde{V}_S^1 größer \tilde{V}_S^2 , die Komponente \tilde{V}_{NR}^1 jedoch kleiner als \tilde{V}_{NR}^2 .

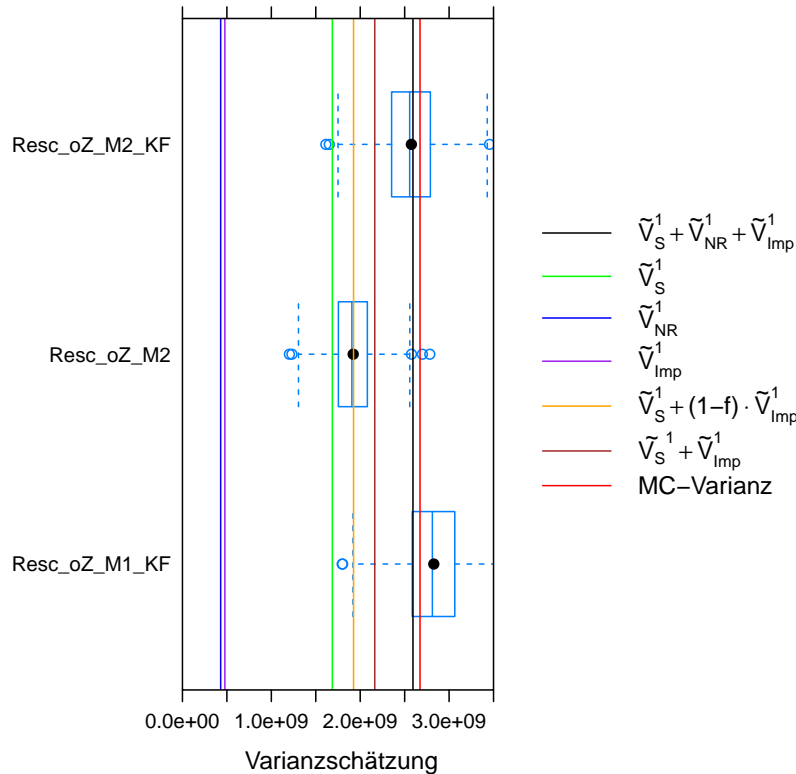


Abbildung 5.33: Abdeckung der einzelnen Varianzkomponenten durch den Rescaling-Bootstrap für einen mittleren Auswahlatz, Hot-Deck-Random-Imputation und Entstehung variabler fehlender Werte in der Grundgesamtheit

In Abbildung 5.33 wird noch einmal gezeigt, welche Versionen des Rescaling-Bootstrap, welche Komponenten der Varianz im Falle variabler fehlender Werte abdecken. So deckt die Version ohne Berücksichtigung des Korrekturfaktors aber mit einer getrennten Adjustierung der für die Substichprobe gezogenen und nicht gezogenen imputierten Werte (*Resc_oZ_M2*) nicht ausschließlich wie im deterministischen Fall die Stichprobenkomponente \tilde{V}_S^1 ab. Diese wird weiterhin durch die grüne Linie repräsentiert und es ist zu sehen, dass diese Version des Rescaling-Bootstrap darüber liegt. Dies macht Sinn, da im Falle zufälliger Imputationsverfahren noch die Imputationskomponente \tilde{V}_{Imp}^1 hinzukommt. Die braune Linie stellt die Summe der Komponenten \tilde{V}_S^1 und \tilde{V}_{Imp}^1 dar. Wie zu sehen ist, liegt die betrachtete Version des Rescaling-Bootstrap jedoch darunter. Und dies entspricht dem in Abschnitt 4.7.1 erläuterten Sachverhalt, wonach bei einem großen Auswahlatz eine schlechte Approximationsgüte der gemeinsamen Verteilung an Stichprobe und Donoren vorliegt. So ist der Rescaling-Bootstrap ohne Korrekturfaktor nur in der Lage den Term $\tilde{V}_S^1 + (1-f) \cdot \tilde{V}_{Imp}^1$ abzudecken und dies wird durch die Abbildung bestätigt. Dieser Term wird durch die orangene Linie repräsentiert und es ist zu sehen, dass die Schätzung dieser Version im Mittel auf dieser Linie liegt. Sollen alle drei Komponenten erfasst werden, muss der Korrekturfaktor im Falle variabler fehlender Werte aus (4.59) berücksichtigt werden.

Dies geschieht beim Schätzer *Resc_oZ_M2_KF* und es ist festzustellen, dass dieser auf der schwarzen Linie liegt und damit die Summe der drei Varianzkomponenten abdeckt. Gegenüber der Monte-Carlo-Varianz ist zu sehen, dass noch kleinere Ungenauigkeiten durch Zufälligkeiten der Simulation vorliegen. Unter Umständen werden hier auch mehr Wiederholungen benötigt. Wie erwartet liegt der Rescaling-Bootstrap ohne separate Re-imputation der nicht gezogenen Elemente über der Varianz und sollte nicht verwendet werden.

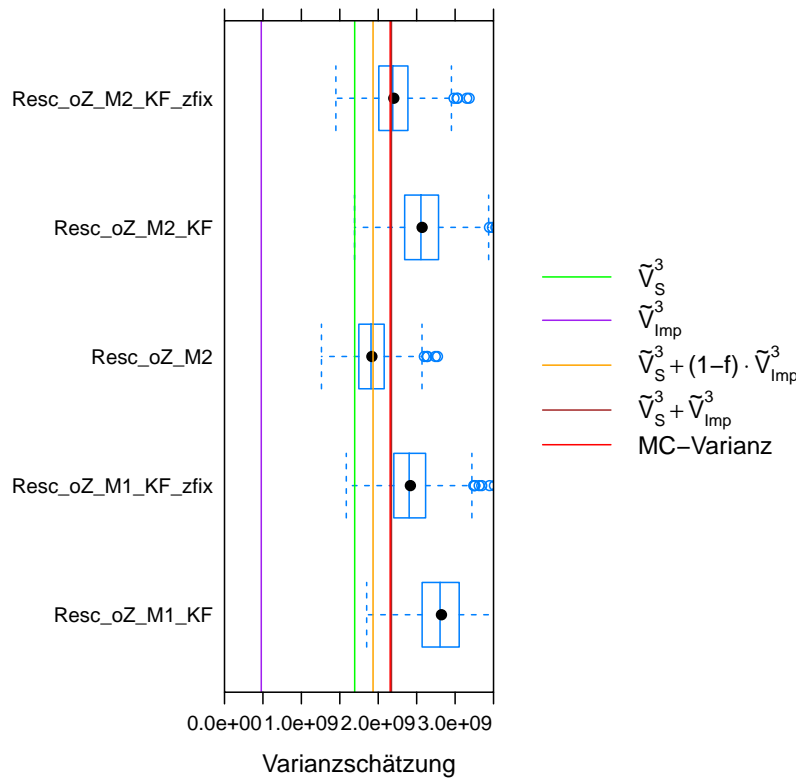


Abbildung 5.34: Abdeckung der einzelnen Varianzkomponenten durch den Rescaling-Bootstrap für einen mittleren Auswahlsatz, Hot-Deck-Random-Imputation und eines fixen Response-Vektors in der Grundgesamtheit

In Abbildung 5.34 wird hingegen der Fall eines fixen Response-Vektors behandelt. Hier gelten die Erläuterungen aus Abschnitt 5.2.4.2, welche aber noch einmal komponentenspezifisch betrachtet werden können. Wie im variablen Fall deckt der Schätzer ohne Korrekturfaktor *Resc_oZ_M2* die Komponente $\tilde{V}_S^3 + (1-f) \cdot \tilde{V}_{Imp}^3$ ab. Im Fall eines fixen Response-Vektors entfällt jedoch die Varianzkomponente bezüglich fehlender Werte \tilde{V}_{NR}^1 . Die Gesamtvarianz setzt sich nur aus den Komponenten $\tilde{V}_S^3 + \tilde{V}_{Imp}^3$ zusammen. Wird dennoch der Korrekturfaktor aus dem Fall Entstehung variabler fehlender Werte in der Grundgesamtheit genommen, damit der Schätzer *Resc_oZ_M2_KF*, wird die Varianz deutlich überschätzt. Die Komponente \tilde{V}_{NR}^1 wird dadurch eingerechnet, obwohl sie überhaupt nicht anfällt. Es darf, wie in 5.2.4.2 erläutert wird, nur derjenige Korrekturfaktor aus (4.65) einbezogen werden, welcher die schlechte Approximationsgüte der gemeinsamen Verteilung aus Stichprobe und Donoren korrigiert. Dies geschieht beim Schätzer *Resc_oZ_M2_KF_zfix*, welcher separat die Adjustierung der imputierten Werte der für Substichprobe gezogenen und nicht gezogenen Elemente vornimmt. Der Schätzer *Resc_oZ_M1_KF_zfix* ohne Adjust-

tierung der imputierten Werte der nicht für die Substichprobe gezogenen Elemente führt auch hier zu deutlichen Überschätzungen.

Vom Fall fehlender Werte in der Stichprobe wird hier abstrahiert. Die relevanten Erkenntnisse wurden im deterministischen Fall gegeben.

5.3.6 Varianzzerlegung für verschiedene Parameter für Hot-Deck-Random-Imputation

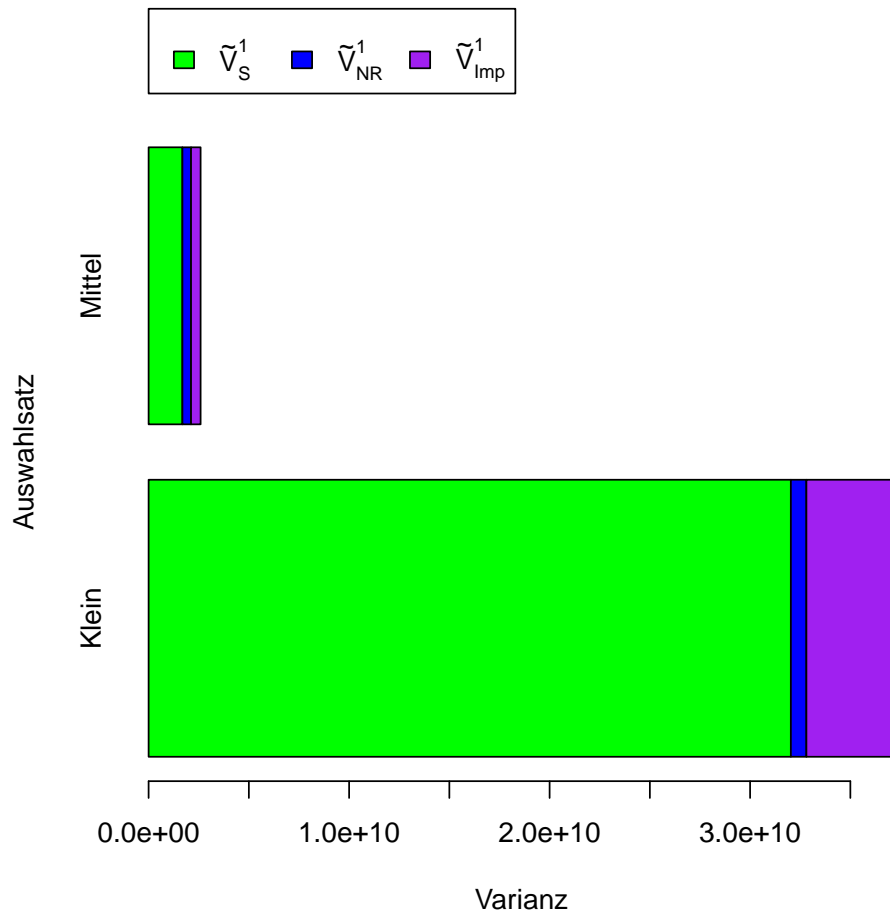


Abbildung 5.35: Varianzzerlegung hinsichtlich verschiedener Auswahlsätze für Hot-Deck-Random-Imputation, Entstehung variabler fehlender Werte in der Grundgesamtheit und mittlerer Auswahlrate

Dieser Abschnitt behandelt die simulative Überprüfung der Veränderung der Varianz und ihrer Komponenten für Hot-Deck-Random-Imputation in Abhängigkeit des Auswahlsatzes und der Ausfallrate. Der theoretische Hintergrund wurde in Abschnitt 3.4.3 gegeben, insbesondere in Bezug auf Formel (3.35). Die dort getroffenen Aussagen können an dieser Stelle grundsätzlich bestätigt werden. Wird zunächst der Fall der Variation des Auswahlsatzes in Abbildung 5.35 betrachtet, kommt für zufällige Imputationsverfahren die Imputationskomponente \tilde{V}_{Imp}^1 hinzu. Es ist zu sehen, dass die Komponente von diesem Parameter abhängt. Es gilt die Aussage aus (3.35), dass \tilde{V}_{Imp}^1 bei kleinem Auswahlwahlsatz groß ist und mit steigendem Auswahlsatz kleiner wird. Für die Stichprobenkomponente \tilde{V}_S^1 und

die Varianzkomponente bezüglich fehlender Werte \tilde{V}_{NR}^1 gelten die Aussagen aus dem deterministischen Fall, da wie in 3.4.3 erwähnt, gegenüber der Mittelwertimputation unter den gegebenen Bedingungen nur die Komponente (3.35) hinzukommt. Die Komponente \tilde{V}_S^1 ist dadurch bei kleinem Auswahlsatz groß und sinkt mit dessen Zunahme. Beim Term \tilde{V}_{NR}^1 sind bezüglich der beiden Ausfallsätze geringe Unterschiede festzustellen. Diese können das Resultat von Zufälligkeiten der Monte-Carlo-Simulation sein. Unter Umständen werden gerade im Fall eines kleineren Auswahlsatzes mehr Wiederholungen benötigt zu einer genaueren Bestimmung der einzelnen Komponenten. Es ist jedoch auch möglich, dass sich kleinere Unterschiede bezüglich beider Ausfallraten in den beiden Szenarien ergeben können. Zwar ist es hier das Ziel, diese Rate zwischen den beiden Szenarien konstant zu halten. Wie bereits in Abschnitt 5.1.4 erläutert, können sich jedoch aufgrund des dort beschriebenen stochastischen Prozesses der Erzeugung fehlender Werte geringfügige Unterschiede in den Ausfallraten ergeben. Diese können unter Umständen zu den Unterschieden in der Komponente \tilde{V}_{NR}^1 führen. Jedoch sind diese Unterschiede, insbesondere im Vergleich zu den beiden anderen Komponenten, zwischen den beiden Auswahlsätzen gering. Da die Komponenten \tilde{V}_S^1 und \tilde{V}_{Imp}^1 mit Zunahme des Auswahlsatzes deutlich sinken, gewinnt die Komponente \tilde{V}_{NR}^1 bei großem Auswahlsatz auch bei Hot-Deck-Random-Imputation an Bedeutung.

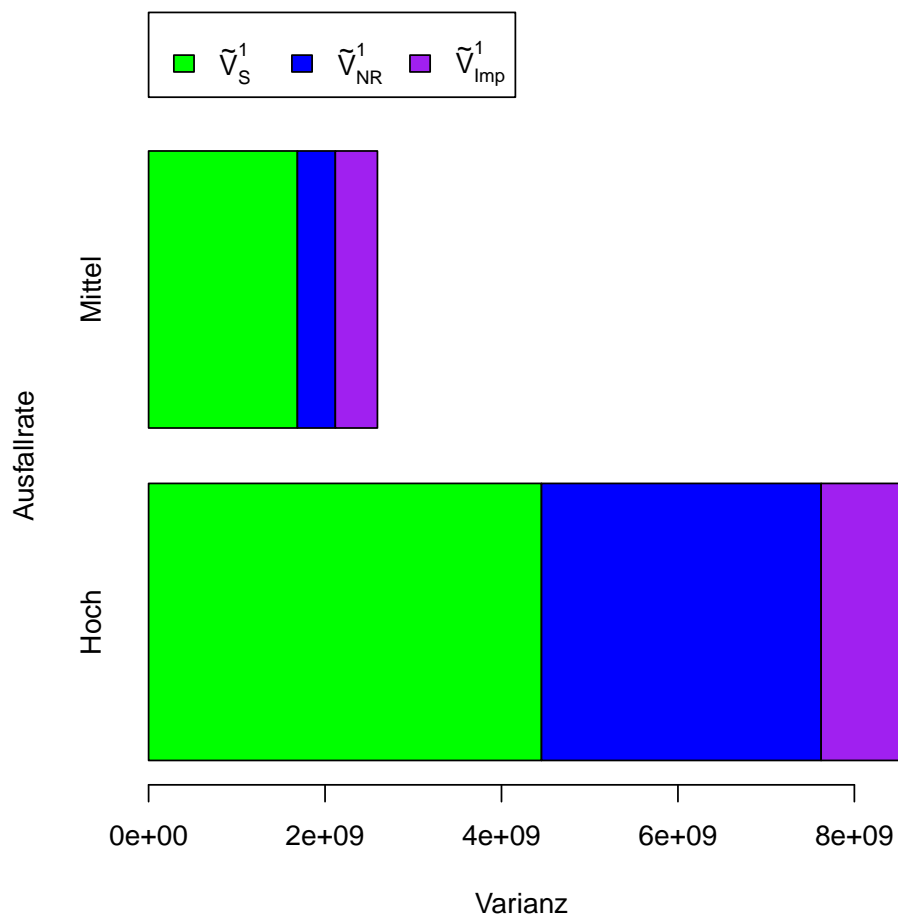


Abbildung 5.36: Varianzzerlegung hinsichtlich verschiedener Ausfallraten für Hot-Deck-Random-Imputation, Entstehung variabler fehlender Werte in der Grundgesamtheit und mittlerem Auswahlsatz

In Abbildung 5.36 wird die Ausfallrate variiert. Auch hier gelten die Aussagen aus Abschnitt 3.4.3, insbesondere Formel (3.35). Es ist zu sehen, dass die Komponente \tilde{V}_{Imp}^1 mit größerer Ausfallrate zunimmt. Dies resultiert aus der größeren Anzahl an fehlenden Werten und die daraus folgende höhere Anzahl an Imputationen. Wie im deterministischen Fall nehmen auch die Komponenten \tilde{V}_S^1 und \tilde{V}_{NR}^1 zu. Damit steigt auch hier die Gesamtvarianz mit größer werdender Ausfallrate unter sonst gleichen Bedingungen deutlich an. Wie an der Abbildung 5.36 zu sehen, verzeichnet naturgemäß die Komponente \tilde{V}_{NR}^1 den größten Anstieg.

Auf die Fälle eines großen Ausfallsatzes bzw. einer kleinen Ausfallrate wird an dieser Stelle verzichtet, aufgrund der deutlich längeren Laufzeiten der Varianzzerlegung für die Hot-Deck-Random-Imputation. Die dargestellten Fälle geben einen ausreichenden Einblick in die Auswirkungen der Änderung der beiden Parameter auf die Varianz und ihrer Komponenten.

5.3.7 Fazit bezüglich der Zerlegung der Gesamtvarianz

In diesem Abschnitt wurden speziell die Ausführungen aus den Abschnitten 3.4.2 und 3.4.3 anhand einer Monte-Carlo-Varianzzerlegung überprüft und bestätigt. Es konnte gezeigt werden, dass Parameter wie der Auswahlatz und die Ausfallrate Auswirkungen auf die absolute Höhe der einzelnen Komponenten wie auch ihres relativen Anteils an der Gesamtvarianz haben. Gleiches gilt für den Ort der Entstehung fehlender Werte, welcher neben der Zusammensetzung der Gesamtvarianz auch für das Auftreten bestimmter Komponenten der Varianz verantwortlich ist. Zu guter Letzt spielt auch die Imputationsmethode eine wichtige Rolle, welche bestimmt, ob die Komponente \tilde{V}_{Imp}^1 auftritt oder nicht.

5.4 Ergebnisse bezüglich der mehrstufigen Stichprobendesigns

Dieser Abschnitt stellt die Simulationsergebnisse der in den Abschnitten 5.1.2 und 5.1.3 beschriebenen mehrstufigen Stichprobendesigns vor. Es wird zunächst ein zweistufiges Stichprobendesign mit einer Stratifizierung auf der ersten Stufe verwendet. Für die Varianzzerlegung wird auf eine Stratifizierung aus den in Abschnitt 5.1.3 genannten Gründen verzichtet. Die Simulationen gehen für diese Designs mit größeren Laufzeiten einher. Dies ist die Konsequenz der höheren Stufenanzahl der Stichprobendesigns und der daraus resultierenden höheren Komplexität der Anwendung der Resampling-Verfahren. Daher wird für diese Szenarien nur noch der Rescaling-Bootstrap ohne Zurücklegen und der BRR berücksichtigt. Es wird des Weiteren nur noch der Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit betrachtet. Die relevanten Erkenntnisse der beiden anderen Fälle wurden im Rahmen der einstufigen Stichprobendesigns gegeben. Der Ausfallmechanismus ist ausschließlich MAR.

Für beide Methoden werden verschiedene Varianten verwendet. Beim Rescaling-Bootstrap leiten sich diese Varianten aus den in Abschnitt 4.5.5 beschriebenen vier Kategorien ab. In diesem Abschnitt wurden zwei Varianten der Kategorienbildung vorgeschlagen. Es wurde theoretisch erörtert, dass Reimputationen bzw. Adjustierungen der imputierten Werte innerhalb der daraus resultierenden Kategorien durchzuführen sind. Die Varianzschätzer, welche sich daraus ergeben, werden mit Varianten des Rescaling-Bootstrap verglichen, bei denen bestimmte Kategorien zusammen gefasst werden oder Reimputationen bzw. Adjustierungen der imputierten Werte bestimmter Kategorien ausgelassen werden. Daraus

ergeben sich verschiedene Varianzschätzer. Diese müssen eindeutig bezeichnet werden. Für den Rescaling-Bootstrap wird weiterhin die Abkürzung *Resc* verwendet. Es werden jedoch zusätzlich die Kategorien K1 bis K4 angehängt, in welchen die Reimputation bzw. Adjustierung zur Anwendung kommt. Eine zwischen den Kategorien getrennte Reimputation bzw. Adjustierung wird mit einem Unterstrich `_` gekennzeichnet. Bei einer Zusammenlegung der Kategorien wird dieser weggelassen. So wird z.B. der Schätzer, welcher die Reimputation bzw. Adjustierung, wie in Abschnitt 4.5.5 vorgeschlagen, getrennt in allen vier Kategorien vornimmt, mit *Resc_K1_K2_K3_K4* bezeichnet.

Liegt jedoch die in Abschnitt 4.5.5 vorgeschlagene Variante vor, welcher eine separate Anwendung der Reimputation bzw. Adjustierung in den Kategorien K1, K3 und der zusammengelegten Kategorie K2 und K4 zugrunde liegt, so erhält der Schätzer die Bezeichnung *Resc_K1_K3_K2K4*. Bei Auslassung der Reimputation bzw. Adjustierung in einer der vier Kategorien wird die Kategorie in der Bezeichnung des Schätzers nicht aufgeführt. Wird zum Beispiel die Reimputation bzw. Adjustierung in den Kategorien K1 bis K3 separat durchgeführt, Kategorie K4 allerdings ausgelassen, so resultiert *Resc_K1_K2_K3*. Bei zusätzlicher Berücksichtigung eines Korrekturfaktors wird die Abkürzung KF angehängt. Beim BRR werden weniger Varianten berücksichtigt. Diese Resampling-Methode wird separat auf den verschiedenen Stufen angewendet. Erfolgt eine Reimputation bzw. Adjustierung auf beiden Stufen so wird der Schätzer mit *BRR_1_2* bezeichnet. Wird eine Reimputation bzw. Adjustierung nur auf der zweiten Stufe durchgeführt, resultiert die Bezeichnung *BRR_2*. Für den BRR muss, wie in Abschnitt 5.1.6 beschrieben, beachtet werden, dass auf der zweiten Stufe des Stichprobendesigns deutlich mehr als 125 Substichproben gezogen werden. Dies hängt mit der beschriebenen Durchführung dieser Methode auf dieser Stufe zusammen.

Im Rahmen der Ergebnisdarstellung wird zunächst von einer gleichen PSU-Größe ausgegangen. Hier werden die Ergebnisse kleinerer und größerer Auswahlätze getrennt erörtert. Anschließend erfolgt die Betrachtung einer ungleichen PSU-Größe. Die Ergebnisse für Regressionsimputation für ausgewählte Schätzer finden sich im Anhang A in Abschnitt A.3.3. Wie im einstufigen Design findet eine Visualisierung durch Boxplots statt. Die Ergebnisdarstellung bezüglich des mehrstufigen Stichprobendesigns schließt mit der simulativen Zerlegung der Gesamtvarianz für die zweistufige Zufallsstichprobe ohne Schichtung.

5.4.1 Kleine und mittlere Auswahlätze

5.4.1.1 Mittelwertimputation

Dieser Abschnitt dient insbesondere der Überprüfung der in der Arbeit in Abschnitt 4.5.5 vorgeschlagenen Erweiterung des Rescaling-Bootstrap für mehrstufige Stichprobendesigns auf den Fall fehlender Werte und anschließender Imputation. Die Abbildungen 5.37 und 5.38 zeigen die Ergebnisse der Schätzungen für das mehrstufige Design im Falle der Mittelwertimputation und kleinerer³² bzw. mittlerer Auswahlätze.³³ Des Weiteren liegen in beiden Szenarien variable fehlende Werte in der Grundgesamtheit vor, wobei die Varianzkomponente V_{NR}^1 bezüglich fehlender Werte aufgrund des geringeren Gesamtauswahlsatzes f in beiden Szenarien nur eine untergeordnete Rolle spielt.

Die Kernthese ist, dass die Adjustierung bzw. die Reimputation separat in den in Abschnitt 4.5.5 beschriebenen Kategorien durchgeführt werden muss. Auf diesem Weg kann

³²Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.11 im Anhang A.

³³Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.12 im Anhang A.

berücksichtigt werden, dass Imputationen untergeordneter Einheiten auch die Varianz übergeordneter Stufen beeinflussen können. Dies macht es notwendig, zum einen die fehlenden Werte zu reimputieren, welche auf der ersten Stufe gezogen worden sind, unabhängig von denjenigen, welche auf der zweiten Stufe gezogen worden sind. Zum anderen müssen die Reimputationen der imputierten Werte wieder getrennt bezüglich den für die Substichprobe gezogenen und nicht gezogenen fehlenden Werten durchgeführt werden. Wie an den Ergebnissen in Abbildung 5.37 zu sehen ist, führt der Rescaling-Bootstrap ohne Zurücklegen zu den besten und unverzerrten Ergebnissen, wenn er gemäß der in Abschnitt 4.5.5 vorgeschlagenen beiden Varianten in jeder Kategorie separat reimputiert wird. Dies sind die Schätzer $Resc_K1_K2_K3_K4$ bzw. $Resc_K1_K3_K2K4$, wobei bei Letzterem eine Zusammenlegung der Kategorien K2 und K4 stattfindet, zur Ausnutzung des Substichprobenziehungsprozesses beim Rescaling-Bootstrap. Beide Schätzer sind in diesem Fall nahezu unverzerrt. Für die Variante $Resc_K1_K3_K2K4$ treten jedoch in diesem Szenario mehr Ausreißer auf. Wie in Abschnitt 4.5.5 erläutert, kann die Bildung gleicher Kategoriegrößen wie beim Schätzer $Resc_K1_K2_K3_K4$ zu einer stabileren Varianzschätzung führen.

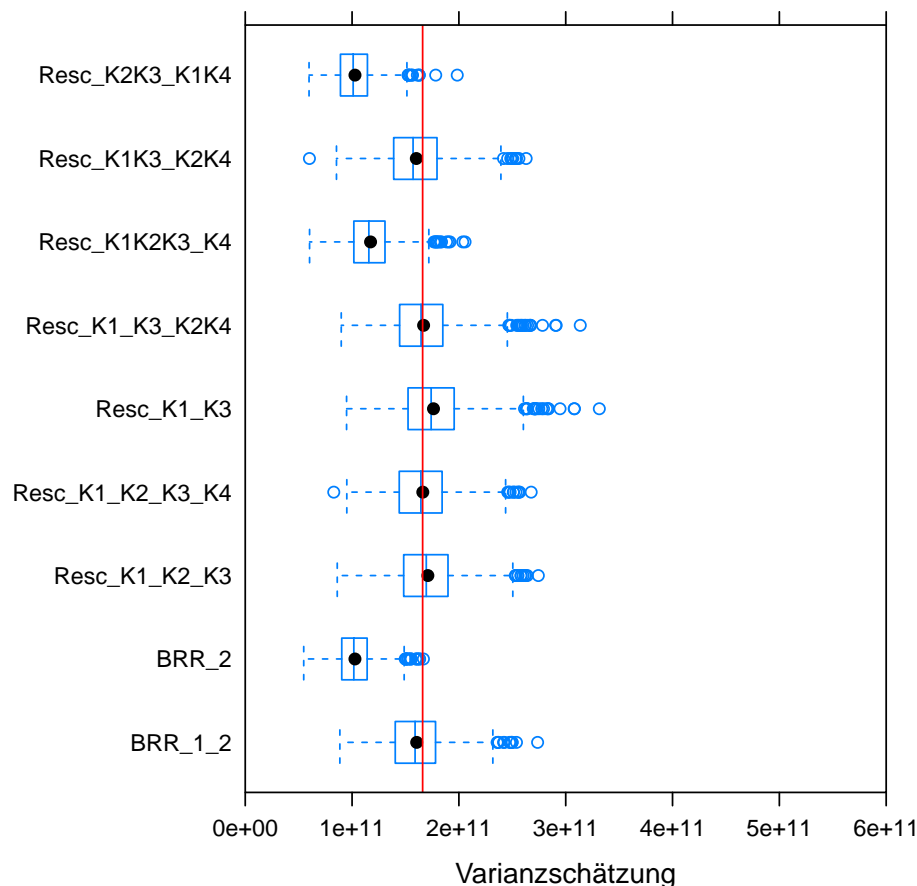


Abbildung 5.37: Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, kleinen Auswahlätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

Des Weiteren lassen sich hier erste Anzeichen einer Überschätzung der Varianz bei Auslassung der Reimputation einer bestimmten Kategorie feststellen. Dies betrifft die Schätzer

Resc_K1_K3 und *Resc_K1_K2_K3*, bei welchen die für die Substichprobe nicht gezogenen Elemente nicht reimputiert wurden. Bei Ersterem wurde dieser Vorgang für die Kategorien 2 und 4 ausgelassen, bei Letzterem nur in Kategorie 4. So zeigt sich für den Schätzer *Resc_K1_K3* in Abbildung 5.37 eine etwas höhere Überschätzung der Varianz. Dies steht im Zusammenhang mit der in dieser Arbeit häufig beschriebenen Variation, welche über die Substichproben dadurch entsteht, dass für eine fehlende Einheit in Abhängigkeit der Substichprobenziehung der imputierte oder reimputierte Wert verwendet wird. Dadurch wird der ursprünglichen Stichproben- und Imputationsprozess nicht richtig wiedergegeben und es resultieren Verzerrungen. Diese sind für den Schätzer *Resc_K1_K2_K3* marginal, da aufgrund der kleinen Auswahlsätze die für die Substichproben nicht gezogenen Elemente in Kategorie K4 nur einen geringen Einfluss besitzen. Dieser Sachverhalt wird im Rahmen großer Auswahlsätze noch einmal aufgegriffen.

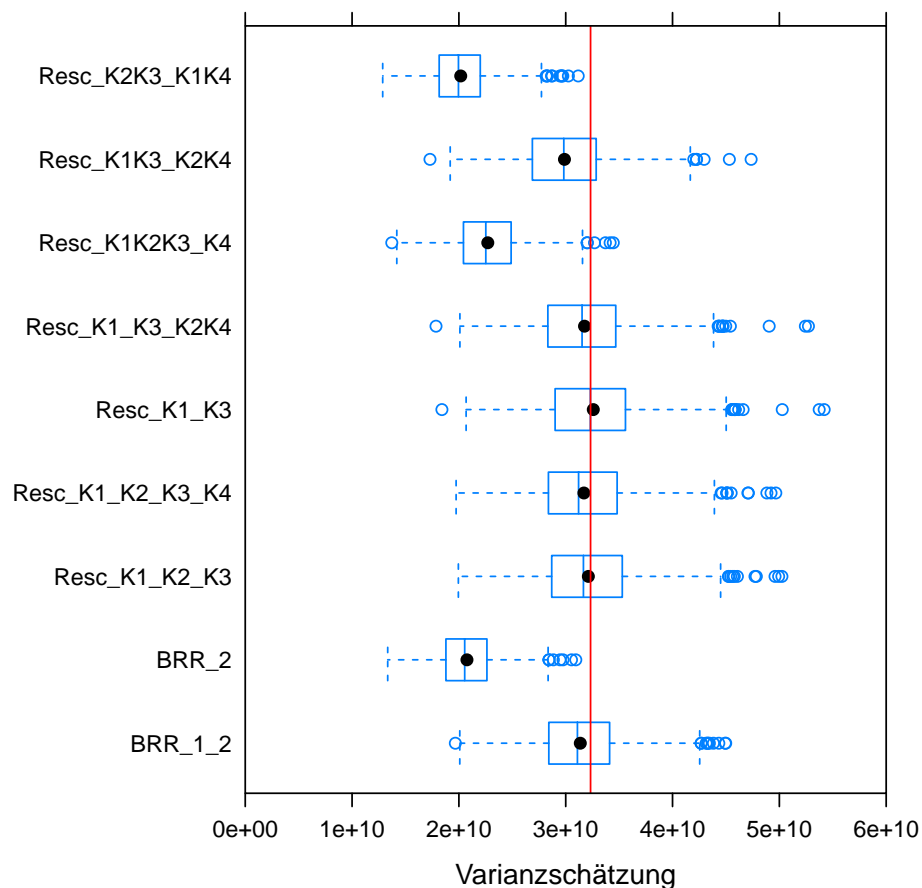


Abbildung 5.38: Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, mittleren Auswahlsätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

Es ist jedoch auch bei den hier vorliegenden Auswahlsätzen gut zu sehen, dass eine separate Reimputation in den einzelnen Kategorien unabdingbar ist. Insbesondere an Abbildung 5.38 wird deutlich, dass die Bildung der Kategorien, wie in Abschnitt 4.5.5 beschrieben, große Vorteile besitzt. Ansonsten wird dem Umstand nicht Rechnung getragen, dass Imputationen untergeordneter Einheiten Auswirkungen auf die Varianz aller Stufen besitzen. Dies zeigen die Schätzer *Resc_K1K3_K2K4*, *Resc_K2K3_K1K4* und *Resc_K1K2K3_K4*. Im

Fälle von *Resc_K1K3_K2K4* wird als Abgrenzung zur Durchführung der Reimputation lediglich die erste Stufe verwendet. Kategorien werden dahingehend abgegrenzt, ob eine Einheit über ihre PSU auf der ersten Stufe in die Substichprobe gelangt ($\delta_{hd} = 1$) oder nicht ($\delta_{hd} = 0$). Für die Reimputation wird durch die Zusammenfassung der Kategorien K1 und K3 nicht berücksichtigt, ob eine Einheit zusätzlich auf der zweiten Stufe gezogen worden ist. Der Einfluss der Imputation auf die Varianz der zweiten Stufe wird dadurch nicht korrekt wiedergegeben. Beim Schätzer *Resc_K2K3_K1K4* verhält es sich gegenteilig. Hier wird im Rahmen der Reimputation nur betrachtet, ob eine Einheit auf der zweiten Stufe für die Substichprobe gezogen worden ist ($\delta_{hdi} = 1$) oder nicht ($\delta_{hdi} = 0$). Die Auswirkungen der Imputation auf die Varianz der ersten Stufe werden vernachlässigt. In beiden Fällen wird dadurch die Varianz unterschätzt, weil die Varianz auf den einzelnen Stufen nicht korrekt erfasst wird. Während die Unterschätzung für den Schätzer *Resc_K2K3_K1K4* bei beiden Auswahlätzen sehr drastisch ist, tritt diese Verzerrung für Schätzer *Resc_K1K3_K2K4* insbesondere beim mittleren Auswahlatz auf. Auch die Gruppenbildung für den Schätzer *Resc_K1K2K3_K4* spiegelt die Auswirkung der Imputationen auf die Varianz beider Stufen nicht richtig wieder und es treten Unterschätzungen auf. Die Verzerrungen für die drei Schätzer sind deutlich und hängen mit den beschriebenen Sachverhalten zusammen. Kleinere Unterschätzungen sind jedoch im Fall eines mittleren Auswahlsatzes auch denkbar, durch den im Vergleich zum Szenario mit kleineren Auswahlätzen größeren Anteil der Varianzkomponente bezüglich fehlender Werte V_{NR}^1 . Alle Verfahren in Abbildung 5.38 verwenden keinen Korrekturfaktor zur Kompensation dieser Varianzkomponente. Jedoch sind die dadurch entstehenden Unterschätzungen sehr gering, wie z.B. die Schätzer *Resc_K1_K2_K3_K4* bzw. *Resc_K1_K3_K2K4* zeigen. Für die Schätzer *Resc_K1_K2_K3* und *Resc_K1_K3* liegt in diesem Fall der glückliche Umstand vor, dass sich die Überschätzungen durch die fehlenden Reimputationen der nicht für die Substichprobe gezogenen Einheiten und die Unterschätzungen durch den fehlenden Korrekturfaktor gerade ausgleichen.

Die bisherigen Ausführungen dieses Abschnittes betreffen den Rescaling-Bootstrap ohne Zurücklegen, aber die Betrachtung der Abbildungen 5.37 und 5.38 zeigt, dass sie auch für den BRR gelten. Auch hier muss eine Trennung der Reimputation bezüglich der beiden Stufen erfolgen. Im Falle des Schätzers *BRR_1_2* wird eine separate Anwendung auf beiden Stufen und damit auch eine getrennte Reimputation verwendet. Auch eine Abgrenzung bezüglich der für die Substichprobe gezogenen und nicht gezogenen Elemente wird auf beiden Stufen berücksichtigt. Es ist zu sehen, dass dieses Vorgehen auch beim BRR fast unverzerrte Schätzungen mit marginalen Unterschätzungen liefert. Im Gegensatz dazu führt der Schätzer *BRR_2*, welcher eine Reimputation auf der ersten Stufe vernachlässigt, zu drastischen Unterschätzungen. Dies entspricht der in Abschnitt 4.1.3 aufgestellten Theorie, dass zur Erfassung aller Varianzkomponenten zusätzlich Reimputationen auf der ersten Stufe durchgeführt werden müssen, da Imputationen auf der zweiten Stufe des ursprünglichen Imputationsprozesses auch Einfluss auf die Varianz der übergeordneten Stufe haben. Daher sollten bei mehrstufigen Designs und deterministischen Imputationsverfahren Reimputationen im Zusammenhang mit dem BRR auf allen Stufen getrennt durchgeführt werden. Aus laufezeittechnischen Gründen wird im Folgenden auf die Angabe von Szenarien mit mittlerem Auswahlatz verzichtet. Die notwendigen Erkenntnisse ergeben sich aus Szenarien mit kleineren bzw. größeren Auswahlätzen.

5.4.1.2 Hot-Deck-Random-Imputation

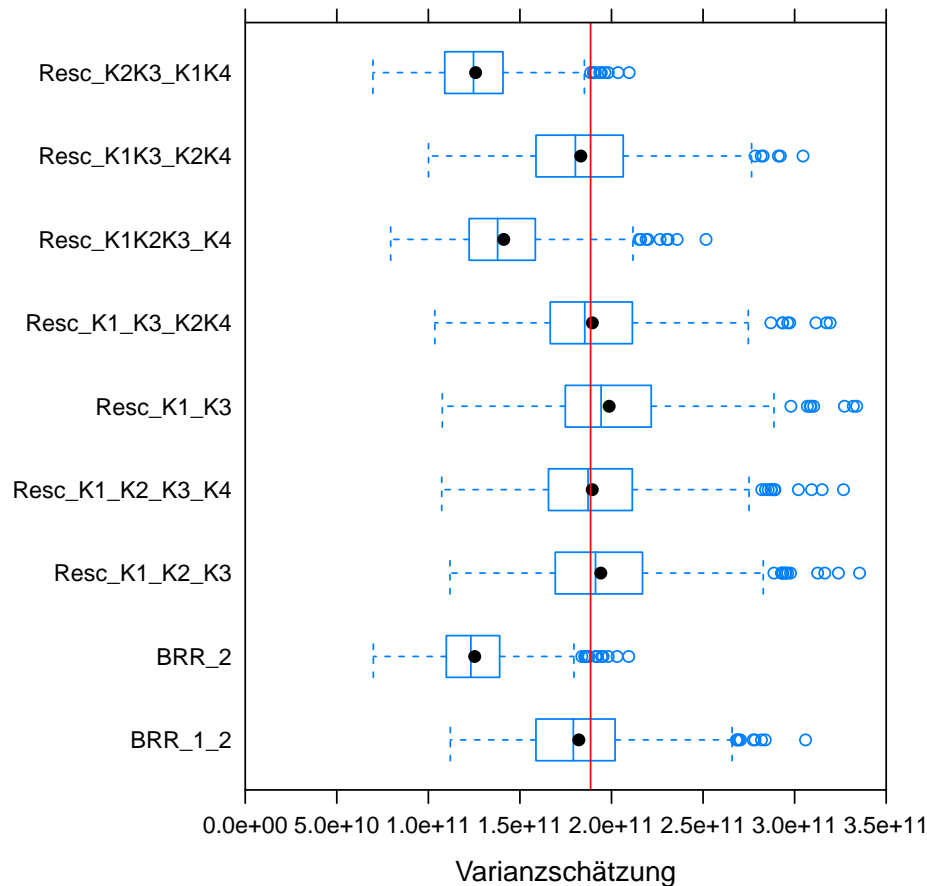


Abbildung 5.39: Vergleich beim zweistufigen Stichprobendesign, Hot-Deck-Random-Imputation, kleinen Auswahlätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

Abbildung 5.39 gibt die Schätzungen im Falle variabler fehlender Werte in der Grundgesamtheit im Rahmen kleiner Auswahlätze auf den verschiedenen Stufen wieder.³⁴ Hier zeigen sich die Ergebnisse des deterministischen Falls aus dem vorherigen Abschnitt. Die Schätzer *Resc_K1_K2_K3_K4* und *Resc_K1_K3_K2K4*, bei welchen die Adjustierungen der imputierten Werte nach den in Abschnitt 4.5.5 beschriebenen Kategorien separat durchgeführt werden, resultieren unverzerrte Schätzungen. Bei den Schätzern ohne Adjustierung der imputierten Werten in manchen Kategorien (*Resc_K1_K2_K3* und *Resc_K1_K3*) zeigen sich erste Anzeichen von Überschätzungen. Da hier vor allem die nicht für die Substichprobe gezogenen Elemente betroffen sind und zusätzlich kleine Auswahlätze vorliegen, sind die Überschätzungen noch nicht übermäßig hoch.

Ansonsten entstehen, wie für die Mittelwertimputation, bei den Schätzern *Resc_K1K3-K2K4*, *Resc_K2K3-K1K4* und *Resc_K1K2K3-K4* Unterschätzungen. Es gelten die Interpretationen aus dem deterministischen Fall. Auch beim BRR müssen die Adjustierungen

³⁴In diesem Szenario gelten die Parameterkonstellationen der Mittelwertimputation aus Abschnitt A.2.2.11 im Anhang A.

auf den Stufen separat durchgeführt werden. Dies geschieht beim Schätzer *BRR_1_2*. Allerdings liegen auch hier kleinere Unterschätzungen vor, welche mit der komplexen Anwendung des BRR im mehrstufigen Design in Verbindung gebracht werden kann.

5.4.1.3 Homogene PSU

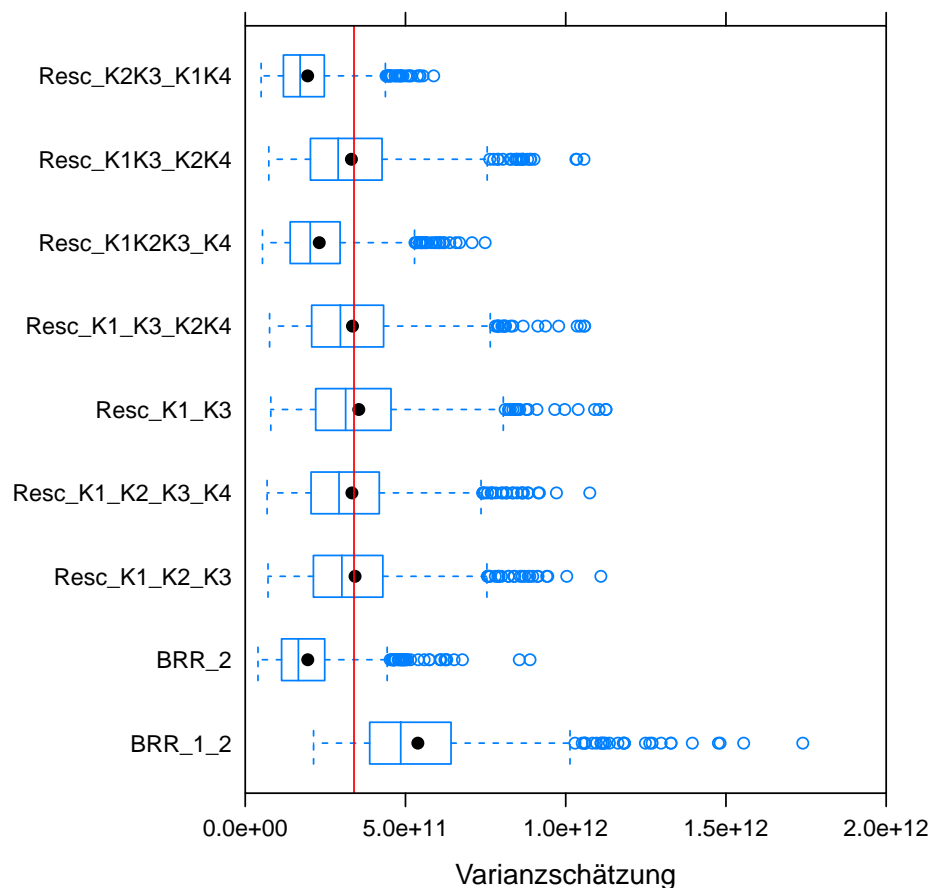


Abbildung 5.40: Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, kleinen Auswahlätzen, homogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

Abbildung 5.40 zeigt die Varianzschätzung für Mittelwertimputation wenn die Untersuchungsvariable dahingehend sortiert wird, dass die PSU sehr homogen sind.³⁵ Dadurch unterscheiden sich die PSU sehr stark und die Varianz entsteht besonders zwischen den Totalwerten der PSU auf der ersten Stufe. Schwierigkeiten für solche Schätzungen sind darin begründet, dass die Schätzer sehr variabel werden können, je nachdem, welche PSU gezogen werden. Dies zeigt sich auch in diesem Beispiel. So führt der Schätzer *BRR_1_2* im Gegensatz zu den Szenarien in Abschnitt 5.4.1.1 zu starken Überschätzungen der Varianz. Der BRR reagiert hier im mehrstufigen Design besonders stark auf die Clusterstruktur. Dies liegt, wie in den Abschnitten 4.6.2 und 4.6.3 beschrieben, an der Gruppierung der PSU im Rahmen der Substichprobenziehung auf der ersten Stufe des BRR. Durch die

³⁵Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.13 im Anhang A.

homogenen PSU wird die Methode sehr variabel und es resultieren in diesem Fall größere Verzerrungen.

Der Rescaling-Bootstrap basierend auf den in Abschnitt 4.5.5 vorgeschlagenen Varianten zur Kategorienbildung für die Reimputation mit den dazugehörigen Schätzern *Resc_K1_K2_K3_K4* bzw. *Resc_K1_K3_K2K4* führt erneut zu fast unverzerrten Ergebnissen. Der Unterschied zum Schätzer *Resc_K1K3_K2K4* ist hier jedoch sehr marginal. Dieser Schätzer vernachlässigt die Auswirkungen der Imputation auf die Varianz der zweiten Stufe. In diesem Fall liegt die Varianz hauptsächlich auf der ersten Stufe. Daher sind die Auswirkungen der Vernachlässigung eher gering.

5.4.2 Große Auswahlsätze

5.4.2.1 Mittelwertimputation

In diesem Szenario wird untersucht, welche Auswirkungen große Auswahlsätze im Zusammenhang mit mehrstufigen Designs besitzen.³⁶ Wie im einstufigen Design werden insbesondere die Berücksichtigung von Korrekturfaktoren, sowie die Reimputation der imputierten Werte der nicht für die Substichprobe gezogenen Elemente überprüft. Dies wird in Abbildung 5.41 demonstriert. Eine fast unverzerrte Schätzung wird nur mit dem Rescaling-Bootstrap ohne Zurücklegen erreicht, welcher eine separate Reimputation in den nach Abschnitt 4.5.5 beschriebenen Kategorien vornimmt und zudem einen Korrekturfaktor zur Berücksichtigung der Varianzkomponente bezüglich fehlender Werte beinhaltet (*Resc_K1_K2_K3_K4_KF* bzw. *Resc_K1_K3_K2K4_KF*). Wird dieser Term nicht integriert, resultiert eine Unterschätzung der Varianz aufgrund der Unterschlagung dieser Varianzkomponente, wie bei den Schätzern *Resc_K1_K2_K3_K4* bzw. *Resc_K1_K3_K2K4*. Wie im einstufigen Fall führt eine Auslassung der Reimputation in bestimmten Kategorien zu einer Überschätzung der Varianz. Dies zeigt sich für die Schätzer *Resc_K1_K2_K3_KF* bzw. *Resc_K1_K3_KF*, bei denen die Imputationswerte der Kategorien K4 bzw. K2 und K4 nicht reimputiert wurden. Ein weiteres Beispiel ist der Schätzer *Resc_K2_K3_KF*, bei dem nur die Elemente der Kategorie K2 und K3 reimputiert wurden. Durch die Auslassung der Reimputation wird wie im einstufigen Design der ursprüngliche Stichproben- und Imputationsprozess nicht richtig abgebildet. Über die verschiedenen Substichproben gehen fehlende Werte, je nachdem, ob sie für die Substichprobe gezogen worden sind oder nicht, mit dem reimputierten Wert oder dem ursprünglichen Imputationswert ein. Dies führt zu mehr Variation als durch den ursprünglichen Imputationsprozess gegeben war. Daher muss bei großen Auswahlätzen zusätzlich auch die Reimputation der Imputationswerte der nicht für die Substichprobe gezogenen Elemente (Kategorie K2 und K4) erfolgen bzw. in allen Kategorien vorgenommen werden. Wie im einstufigen Fall können die nicht für die Substichprobe gezogene Einheiten bei hohen Auswahlätzen ein größeres Gewicht besitzen. Dadurch entstehen größere Überschätzungen im Vergleich zu den Szenarien mit kleineren Auswahlätzen in Abschnitt 5.4.1.1.

³⁶Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.14 im Anhang A. Die Monte-Carlo-Varianz zeigt sich in diesem Szenario weniger stabil. Daher basiert sie auf 50.000 Wiederholungen.

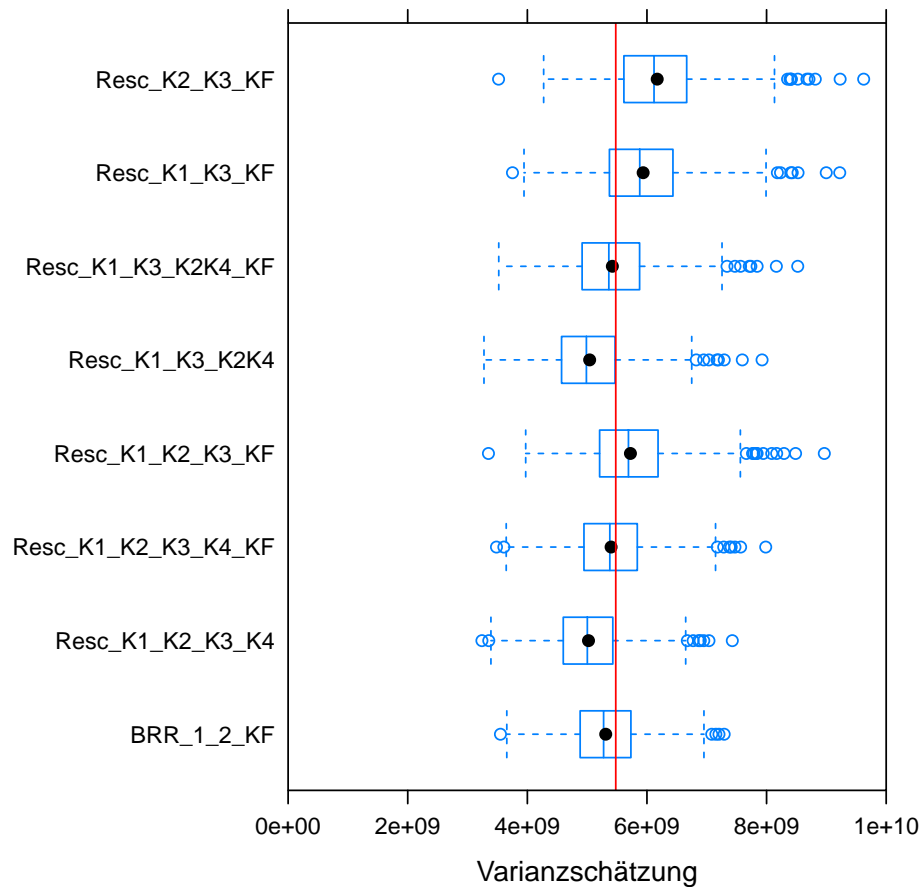


Abbildung 5.41: Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, hohen Auswahlsätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

Die Unterschiede der Simulationsergebnisse der beiden in dieser Arbeit vorgeschlagenen Varianten zur Kategorienbildung zur Reimputation beim Rescaling-Bootstrap (*Resc_K1_K2_K3_K4_KF* und *Resc_K1_K3_K2K4_KF*) in diesem Szenario sind marginal. Es wurde schon beschrieben, dass beide fast unverzerrt sind. Der Schätzer unter Zusammenlegung der Kategorie 2 und 4 liegt etwas näher am Benchmark. Die relative Verzerrung kann gegenüber dem Schätzer *Resc_K1_K2_K3_K4_KF* von 1,4 % auf 1 % reduziert werden. Die Verbesserung darf im Rahmen von Monte-Carlo-Simulationen aber nicht überinterpretiert werden. Der Schätzer *Resc_K1_K3_K2K4_KF* besitzt auch in diesem Szenario eine größere Varianz der Varianz, insbesondere aufgrund der höheren Ausreißer. In Abschnitt 4.5.5 wurde erläutert, dass der Schätzer *Resc_K1_K2_K3_K4_KF* etwas stabiler sein könnte aufgrund der ungefähr gleichen Kategoriengrößen. Dies kann für dieses Szenario bestätigt werden.

Wie bei kleinen Auswahlsätzen wird die Version des BRR (*BRR_1_2_KF*) auf den zwei verschiedenen Stufen separat angewandt, bei einer Reimputation auf beiden Stufen. Zudem wird dieser Vorgang, wie zuvor, getrennt für die Imputationswerte der für die Substichprobe gezogenen und nicht gezogenen Elementen durchgeführt. Zusätzlich verwendet dieses Verfahren den Korrekturfaktor zur Inklusion der Varianzkomponente bezüglich fehlender Werte. Es ist zu sehen, dass auch der auf diesem Wege implementierte BRR zu fast unverzerrten Schätzungen führt. Kleinere Unterschätzungen können mit der Komplexität der

Anwendung des BRR bei mehrstufigen Stichprobendesigns in Zusammenhang stehen.

5.4.2.2 Hot-Deck-Random-Imputation

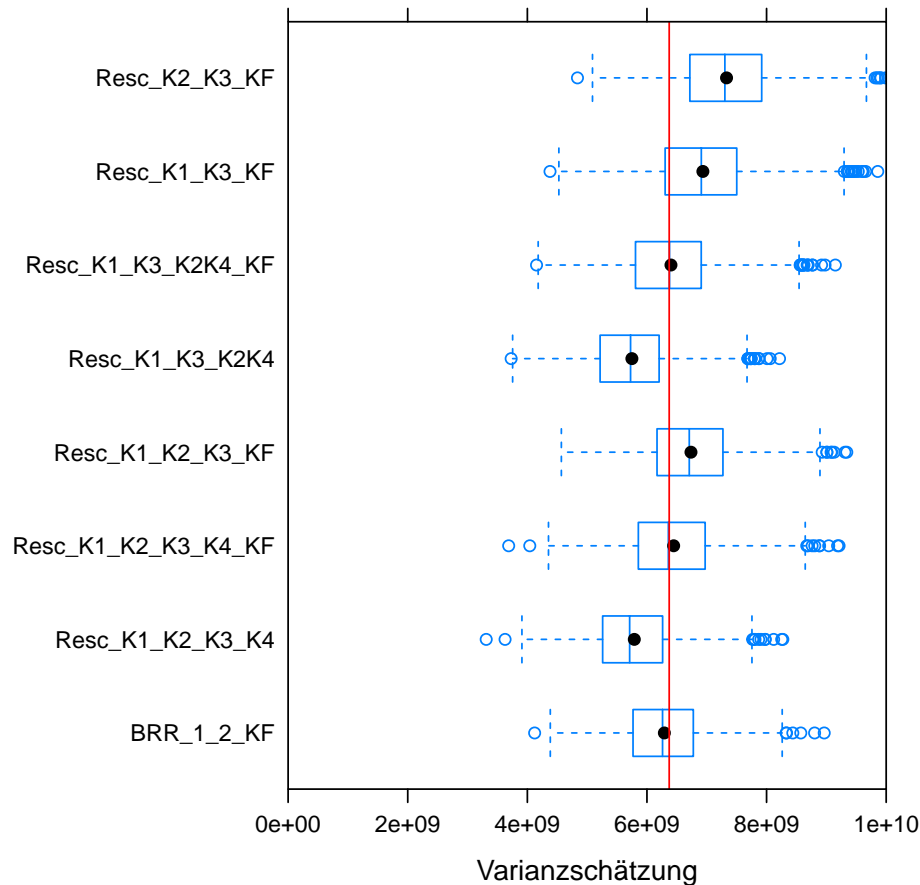


Abbildung 5.42: Vergleich beim zweistufigen Stichprobendesign, Hot-Deck-Random-Imputation, großen Auswahlätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

Auch im Rahmen der Hot-Deck-Random-Imputation in Abbildung 5.42 bei großen Auswahlätzen führen die Schätzer *Resc_K1_K2_K3_K4_KF* und *Resc_K1_K3_K2K4_KF* zu den besten Ergebnissen.³⁷ Verfahren ohne Korrekturfaktor führen zu Unterschätzungen der Varianz. Methoden, welche auf die Adjustierung der imputierten Werte einer Kategorie verzichten, gehen aus den bekannten Gründen mit Überschätzungen einher.

Der Schätzer unter Zusammenlegung der Kategorien K2 und K4 (*Resc_K1_K3_K2K4_KF*) liefert in diesem Fall unverzerrte Schätzungen. Die Anwendung der Adjustierung imputierter Werte separat in vier Kategorien (*Resc_K1_K2_K3_K4_KF*) führt zu einer kleineren Überschätzung. Hier könnte die bessere Ausnutzung des Substichprobenbildungsprozesses des Rescaling-Bootstrap ohne Zurücklegen, wie in Abschnitt 4.5.5 beschrieben, beim Schätzer *Resc_K1_K3_K2K4_KF* von Vorteil sein.

³⁷In diesem Szenario gelten die Parameterkonstellationen der Mittelwertimputation aus Abschnitt A.2.2.14 im Anhang A.

5.4.2.3 Homogene PSU

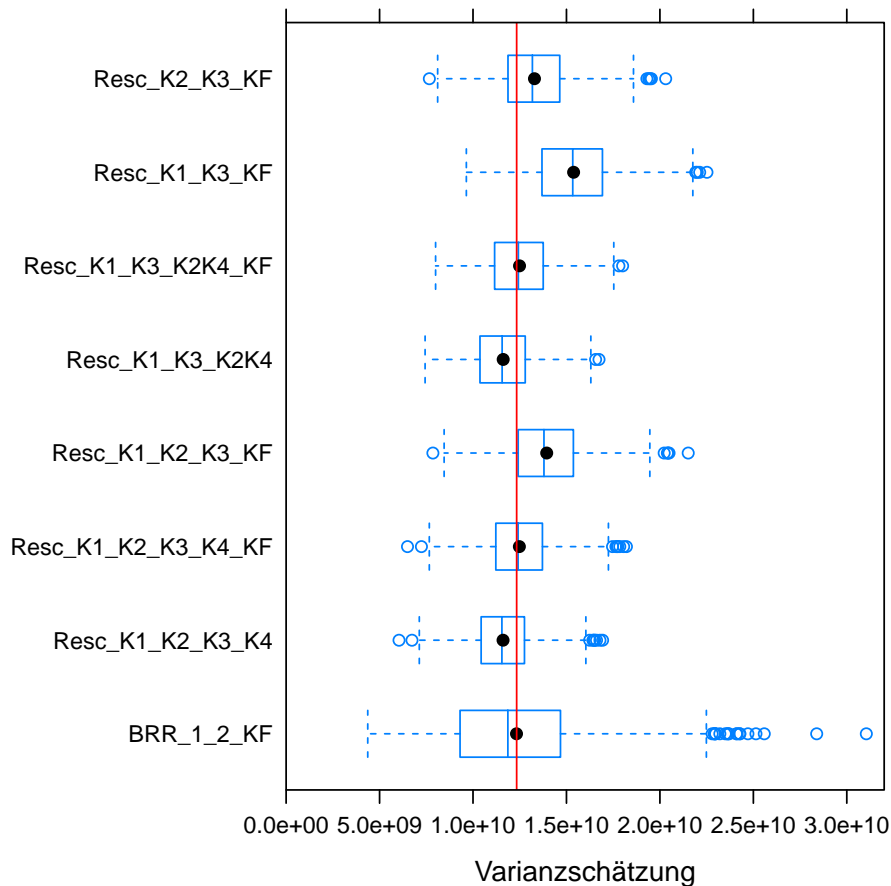


Abbildung 5.43: Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, großen Auswahlätzen, homogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

Die Ergebnisse im Falle großer Auswahlätze, Mittelwertimputation und homogenen PSU sind in Abbildung 5.43 zu sehen.³⁸ Die Schätzer *Resc_K1_K3_K2K4_KF* und *Resc_K1_K2_K3_K4_KF* führen wieder zu unverzerrten Schätzungen. Dies gilt in diesem Fall auch für den BRR. Bei Vorliegen kleinerer Auswahlätze und homogener PSU in Abschnitt 5.4.1.3 wurden für diesen Schätzer noch größere Verzerrungen festgestellt. Jedoch besitzt der BRR im Falle hoher Auswahlätze die deutlich größere Varianz der Varianz als die beiden Schätzer *Resc_K1_K3_K2K4_KF* und *Resc_K1_K2_K3_K4_KF*. Dies steht wieder im Zusammenhang mit der Gruppierung von homogenen PSU auf der ersten Stufe im Rahmen der Substichprobenziehung, wodurch die Schätzungen sehr variabel werden. In Abschnitt 4.7.3.5 wurde erläutert, dass der Korrekturfaktor zur Berücksichtigung der Varianzkomponente bezüglich fehlender Werte V_{NR}^1 bei großen Abweichungen des Stichprobendesigns der einfachen Zufallsstichprobe nicht mehr verwendet werden kann. Als Beispiel wurden homogene PSU angeführt. Dies kann aufgrund der Simulationsergebnisse in Abbildung 5.43 unter den gegebenen Bedingungen nicht bestätigt werden. Hier sind

³⁸Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.15 im Anhang A.

die Abweichungen zur einfachen Zufallsstichprobe noch nicht zu stark ausgeprägt, dass Verzerrungen entstehen. Auch ist es möglich, dass die Komponente V_{NR}^1 aufgrund des hohen Gesamtauswahlsatzes zwar auftritt, aber noch keinen so großen Beitrag besitzt, dass die Verzerrungen aufgrund des komplexen Stichprobendesigns stark ausgeprägt sind.

5.4.3 Vorliegen einer ungleichen PSU-Größe

5.4.3.1 Kleine Auswahlsätze

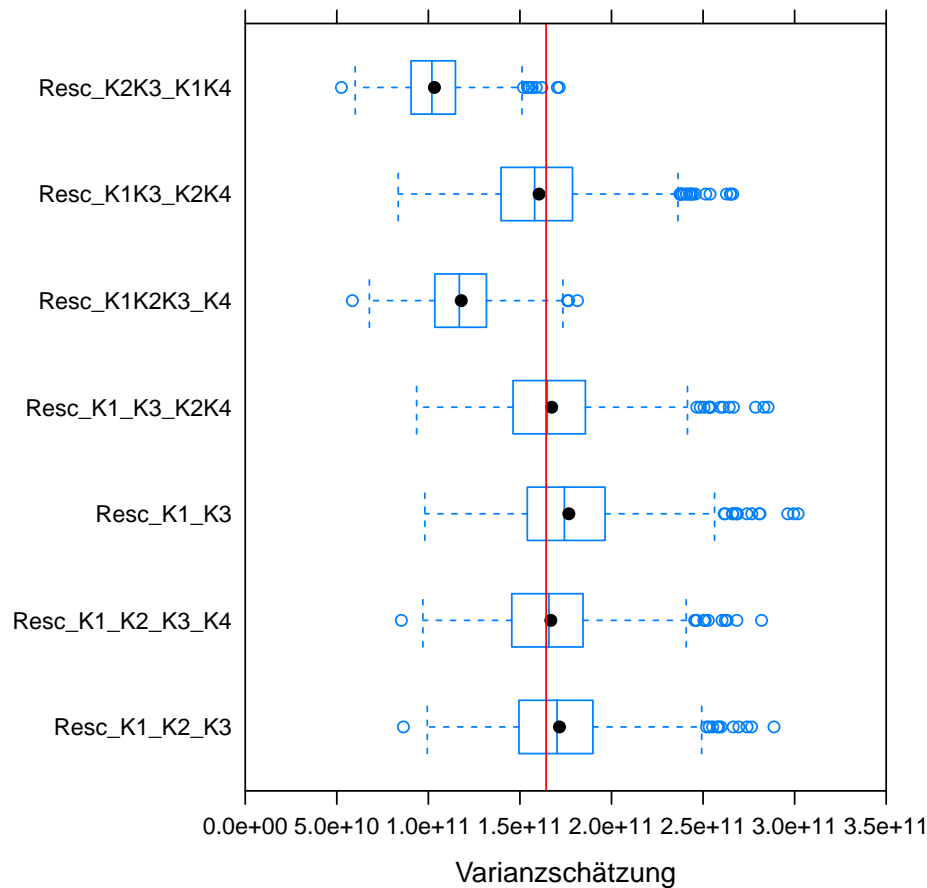


Abbildung 5.44: Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, kleinen Auswahlsätzen, heterogener PSU, einer gering variierenden PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

In den Abbildungen 5.44 und 5.45 wird der Fall einer ungleichen PSU-Größe und kleiner Auswahlsätze behandelt. Die beiden Abbildungen unterscheiden sich in der Variation der PSU-Größe. Außerdem liegen in diesen Szenarien mit über die Schichten variierenden L_h/l_h vor. Damit kann der BRR aus (4.39), wie in Abschnitt 4.6.3 beschrieben, nicht mehr verwendet werden. Die Schätzer $BRR_{1,2}$ und BRR_2 werden daher für die Szenarien mit ungleicher PSU-Größe nicht mehr berücksichtigt. In Abbildung 5.45³⁹ sind die

³⁹Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.17 im Anhang A.

Größenunterschiede stärker ausgeprägt, wohingegen in Abbildung 5.44⁴⁰ sich die bekannten Resultate aus dem Fall einer gleichen PSU-Größe zeigen (vgl. insbesondere Abschnitt 5.4.1.1).

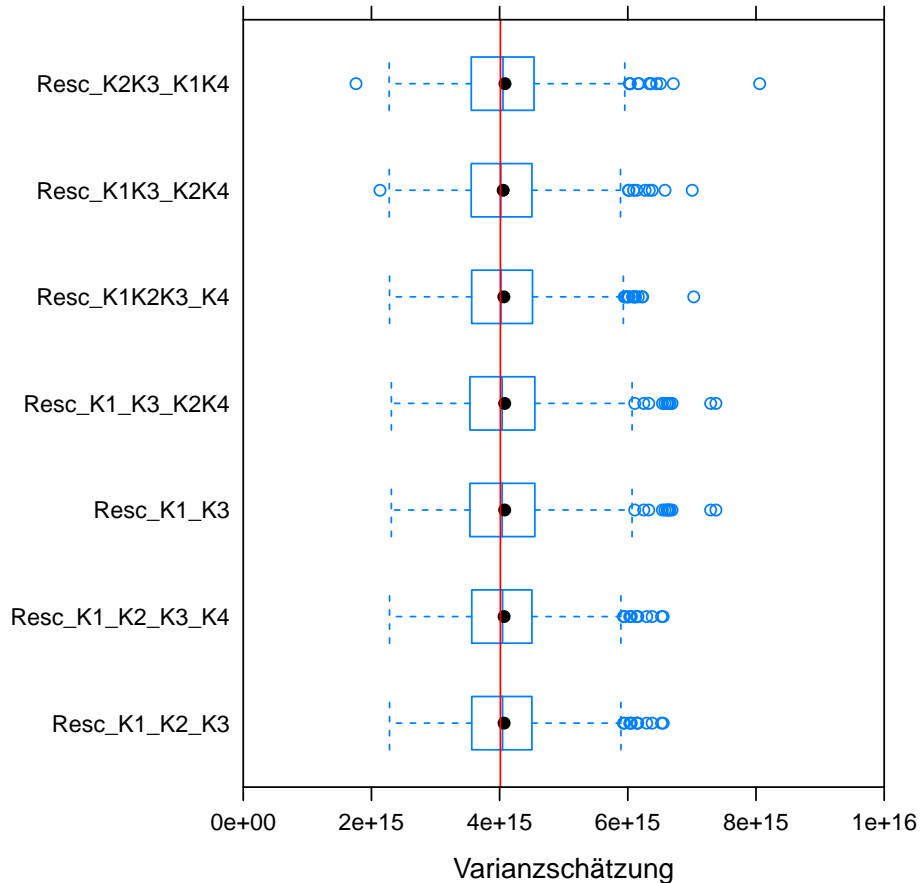


Abbildung 5.45: Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, kleinen Auswahlsätzen, heterogener PSU, einer stark variierenden PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

Die beiden Schätzer $Resc_K1_K2_K3_K4$ bzw. $Resc_K1_K3_K2K4$ führen wieder zu unverzerrten Ergebnissen, während andere Schätzer aus den bekannten Gründen die Varianz über- oder unterschätzen. In Abbildung 5.45 liefern alle Schätzer bei einer stark variierenden PSU-Größe unverzerrte Schätzungen. Die Unterschiede zwischen den einzelnen Schätzern sind gering. Hierfür könnten zwei Ursachen verantwortlich sein. So ist es denkbar, dass durch die unterschiedliche PSU-Größe die durch den Rescaling-Bootstrap angepassten Designgewichte der auf beiden Stufen für die Substichprobe gezogenen Einheiten (Kategorie K3) deutlich wachsen. Diese Gewichte könnten so groß sein, dass die Einheiten der drei anderen Kategorien bei der Schätzung des Rescaling-Punktschätzers keine große Rolle mehr spielen. Des Weiteren wird die Gesamtvarianz deutlich größer. Dies zeigt der Vergleich mit der Abbildung 5.44. Ursache hierfür ist ebenfalls die unterschiedliche PSU-Größe. Diese Varianz schlägt sich insbesondere auf der ersten Stufe des

⁴⁰Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.16 im Anhang A.

Stichprobendesigns nieder. Die Varianzzunahme durch die unterschiedliche PSU-Größe könnte so groß sein, dass andere Varianzkomponenten keinen großen Anteil besitzen. Die dargestellten Resampling-Verfahren scheinen diesen Varianzanteil, welcher durch die unterschiedliche PSU-Größe erzeugt wird, in diesem Fall gut abzudecken. Dies liegt daran, dass die PSU synthetisch erzeugt wurden. Die Größenunterschiede können in der Praxis noch wesentlich ausgeprägter sein und eine Varianzschätzung erschweren.

Vollständig kann dieser Sachverhalt nicht aufgeklärt werden. In Abschnitt 5.4.4.3 wird auch eine Varianzzerlegung im mehrstufigen Fall bei einer ungleichen PSU-Größe durchgeführt. Jedoch ist es hier nicht möglich, die Varianz, welche durch die unterschiedliche PSU-Größe resultiert, von den übrigen Komponenten zu trennen. Dies wäre aber notwendig, um eindeutigere Aussagen zu treffen.

5.4.3.2 Große Auswahlsätze

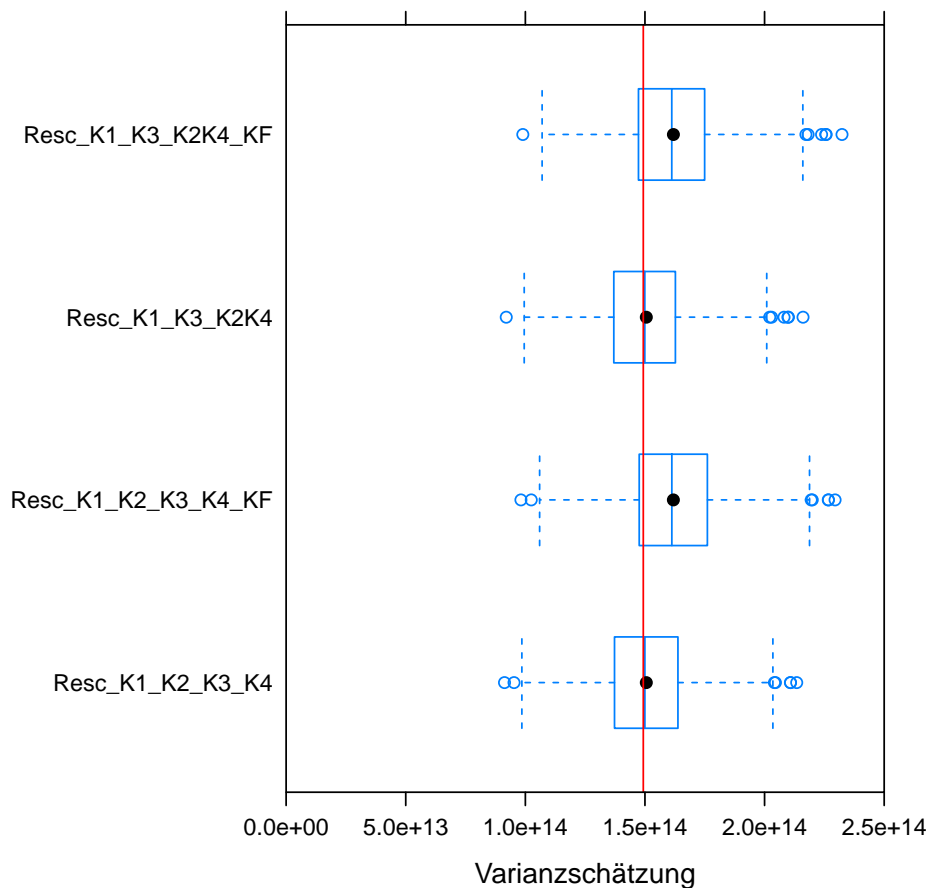


Abbildung 5.46: Vergleich beim zweistufigen Stichprobendesign, Mittelwertimputation, großen Auswahlätzen, heterogener PSU, einer stark variierenden PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

Im Fall großer Auswahlsätze und einer ungleichen PSU-Größe in Abbildung 5.46,⁴¹ zeigt sich ein gegenteiliges Bild zum Fall einer gleichen PSU-Größe aus Abschnitt 5.2.4.1.⁴² Hier führen die Schätzer ohne Korrekturfaktoren *Resc_K1_K2_K3_K4* bzw. *Resc_K1_K3_K2K4* zu unverzerrten Schätzungen. Bei Integration des Korrekturfaktors ergeben sich große Überschätzungen (vgl. *Resc_K1_K2_K3_K4_KF* bzw. *Resc_K1_K3_K2K4_KF*). Durch die unterschiedlichen PSU-Größen könnten sich große Unterschiede zum Fall der einfachen Zufallsstichprobe ergeben. In diesem Fall dürften die Korrekturfaktoren nicht mehr verwendet werden. Die Varianzkomponente bezüglich fehlender Werte ist in diesem Szenario noch nicht so stark ausgeprägt, wodurch die Schätzer *Resc_K1_K2_K3_K4* bzw. *Resc_K1_K3_K2K4* unverzerrt sind. Dies könnte ebenfalls mit der Varianzkomponente bezüglich der unterschiedlichen PSU-Größe in Verbindung stehen. Diese Varianzkomponente könnte in diesem Szenario so groß sein, dass die Varianz bezüglich fehlender Werte vernachlässigbar gering ist.

5.4.4 Varianzzerlegung im mehrstufigen Stichprobendesign ohne Stratifizierung

Für das mehrstufige Stichprobendesign wird ebenfalls eine Varianzzerlegung durchgeführt. Wie in Abschnitt 5.1.3 erläutert, sind hier die Simulationen deutlich aufwendiger. Daher wird die Varianzzerlegung für wesentlich weniger Parameter und Szenarien durchgeführt. Es werden vier Szenarien berücksichtigt. Die Varianzzerlegung wird zunächst im Rahmen kleiner Auswahlsätze und einer gleichen PSU-Größe für heterogene und homogene PSU betrachtet. Anschließend erfolgt die Ergebnisdarstellung für kleine Auswahlsätze, heterogene PSU und einer ungleichen PSU-Größe. Die Interpretation der Simulationsergebnisse schließt mit der Betrachtung des Falls großer Auswahlsätze, heterogener PSU und einer gleichen PSU-Größe. Alle Szenarien beziehen sich auf den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit und der Anwendung von Mittelwertimputation.

5.4.4.1 Varianzzerlegung für den Fall heterogener PSU, einer gleichen PSU-Größe und kleiner Auswahlsätze

Die Varianzzerlegung für den Fall heterogener PSU, einer gleichen PSU-Größe und kleiner Auswahlsätze ist in Abbildung 5.47 dargestellt.⁴³ Der obere rote Balken repräsentiert die Monte-Carlo-Varianz, als Benchmark für die gesamte Varianz. Darunter finden sich zwei verschiedene Varianzzerlegungen. Die Varianzzerlegung *VZ_A* verwendet für die Varianz der ersten Stufe diejenige Varianz dieser Stufe, welche im Falle vollständiger Beobachtungen vorliegt. Diese Komponente wird mit V_1 bezeichnet. Die Varianzzerlegung *VZ_B* verwendet die durch die Imputation niedrigerer Stufe beeinflusste Varianz der ersten Stufe $V_{S_1}^1$. Beide Varianzen der ersten Stufe sind in der jeweiligen Varianzzerlegung durch den gelben Balken dargestellt. Durch die Verwendung der beiden Varianzzerlegungen sollen die Auswirkungen der Imputation auf die Varianz der übergeordneten Stufe gezeigt werden. Wie zu sehen ist, nimmt der gelbe Balken in *VZ_B* aufgrund der Imputationen zu. Ein

⁴¹Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.18 im Anhang A.

⁴²Gegenüber der Abbildung 5.41 werden aus Gründen der längeren Laufzeiten und des darzustellenden Sachverhaltes nur noch die Schätzer *Resc_K1_K2_K3_K4*, *Resc_K1_K3_K2K4* bzw. *Resc_K1_K2_K3_K4_KF* bzw. *Resc_K1_K3_K2K4_KF* berücksichtigt. Es wird außerdem nur der Fall einer stark variierenden PSU-Größe betrachtet.

⁴³Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.19 im Anhang A.

Vergleich mit dem Benchmark zeigt, dass nur die Varianzzerlegung VZ_B die Gesamtvarianz richtig darstellt. Die Gesamtvarianz in VZ_A ist zu niedrig, da die Imputation der zweiten Stufe Einfluss auf die Varianz der ersten Stufe besitzt. Dieser Sachverhalt wurde in der Arbeit grundlegend angenommen. Die Varianzkomponente V_{S1}^1 und V_{S2}^1 bilden zusammen die Stichprobenkomponente V_S^1 . In diesem Fall ist die Komponente V_{S2}^1 , welche sich auf die zweite Stufe des Stichprobendesigns bezieht und repräsentiert wird durch den grünen Balken, deutlich größer. Da die PSU sehr heterogen sind, liegt mehr Variabilität bezüglich der Stichprobenziehung auf der zweiten Stufe vor. Daher ist diese Komponente größer. Die Varianzkomponente bezüglich fehlender Werte V_{NR}^1 besteht nur aus der Komponente V_{NR2}^1 , dargestellt durch den blauen Balken. Die Ursache hierfür ist, dass fehlende Werte nur auf der letzten Stufe erzeugt worden sind. Auf der ersten Stufe tritt diese Komponente nicht in Erscheinung. Aber auch die Komponente V_{NR2}^1 besitzt nur einen sehr geringen Anteil. Der Grund hierfür ist der geringe Gesamtauswahlsatz.

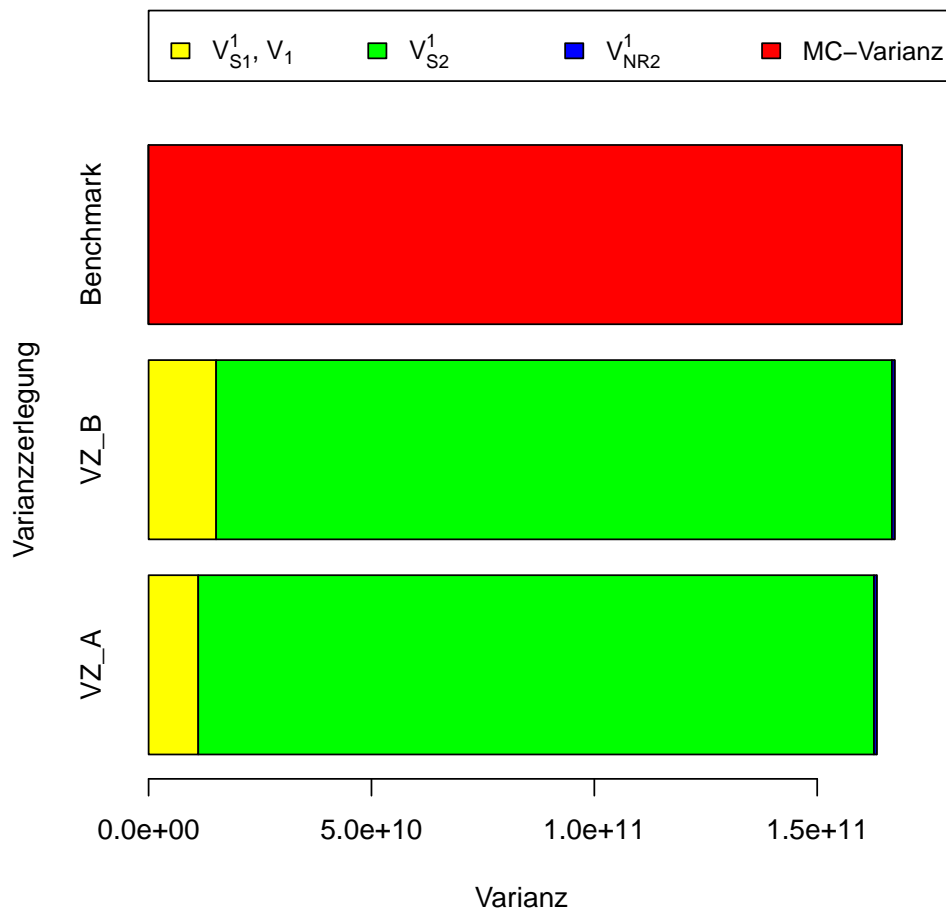


Abbildung 5.47: Varianzzerlegung für den Fall kleiner Auswahlsätze, heterogener PSU und einer gleichen PSU-Größe

5.4.4.2 Varianzzerlegung für den Fall homogener PSU, einer gleichen PSU-Größe und kleiner Auswahlsätze

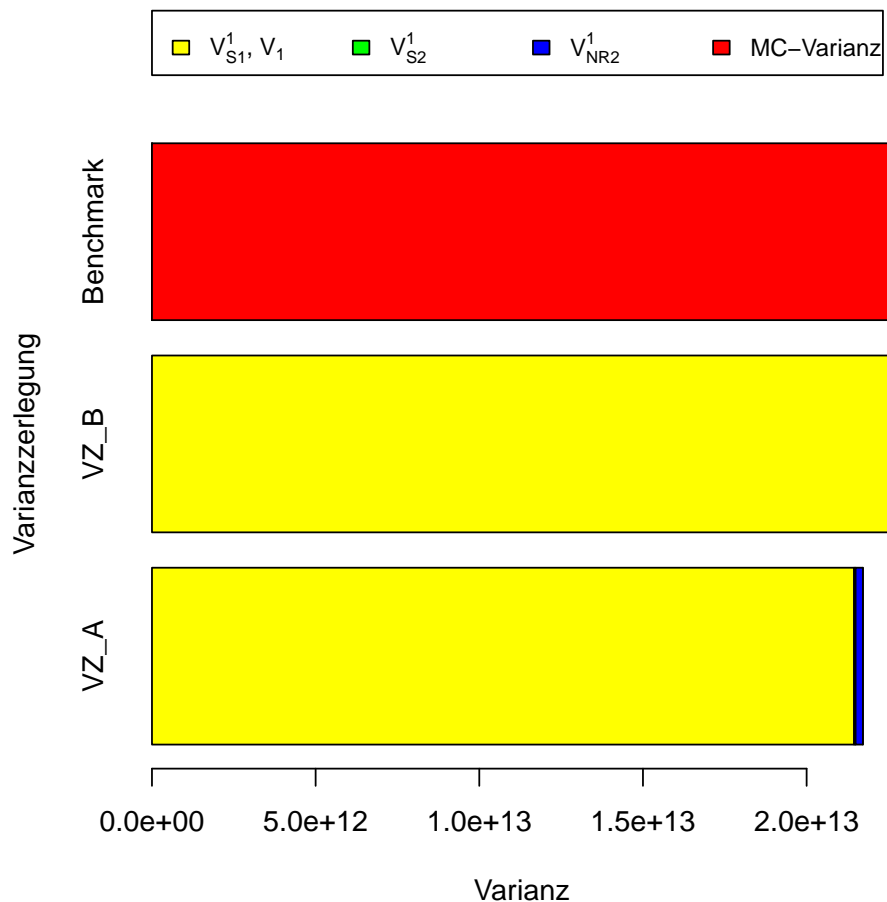


Abbildung 5.48: Varianzzerlegung für den Fall kleiner Auswahlsätze, homogener PSU und einer gleichen PSU-Größe

Auch im Falle homogener PSU in Abbildung 5.48 ist beim Vergleich der Varianzzerlegungen VZ_A und VZ_B zu erkennen, dass die Varianz der untergeordneten Stufe die Varianz der ersten Stufe beeinflusst.⁴⁴ Nur die Varianzzerlegung VZ_B beschreibt die Gesamtvarianz korrekt. In diesem Fall ist jedoch die Stichprobenkomponente V_{S1}^1 bezüglich der ersten Stufe deutlich größer. Dies liegt daran, dass die PSU untereinander sehr verschieden sind, jedoch die Varianz innerhalb der PSU gering ist. Dadurch resultiert mehr Varianz auf der ersten Stufe. Die Komponente V_{NR2}^1 ist wieder vernachlässigbar, aufgrund des geringen Gesamtauswahlsatzes.

5.4.4.3 Varianzzerlegung für den Fall heterogener PSU, einer unterschiedlichen PSU-Größe und kleiner Auswahlsätze

Der Fall einer ungleichen PSU-Größe ist dem Fall von homogenen PSU sehr ähnlich. Zwar sind die PSU in diesem Szenario heterogen.⁴⁵ Durch die ungleiche PSU-Größe resultiert

⁴⁴Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.20 im Anhang A.

⁴⁵Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.21 im Anhang A.

jedoch viel Varianz auf der ersten Stufe. Dies zeigt Abbildung 5.49. Die Stichprobenkomponente bezüglich der ersten Stufe V_{S1}^1 dominiert hier die Varianz. Die Komponente V_{S2}^1 ist sehr klein. Auch die Komponente V_{NR2}^1 besitzt nur einen geringen Anteil wegen des geringen Auswahlsatzes.

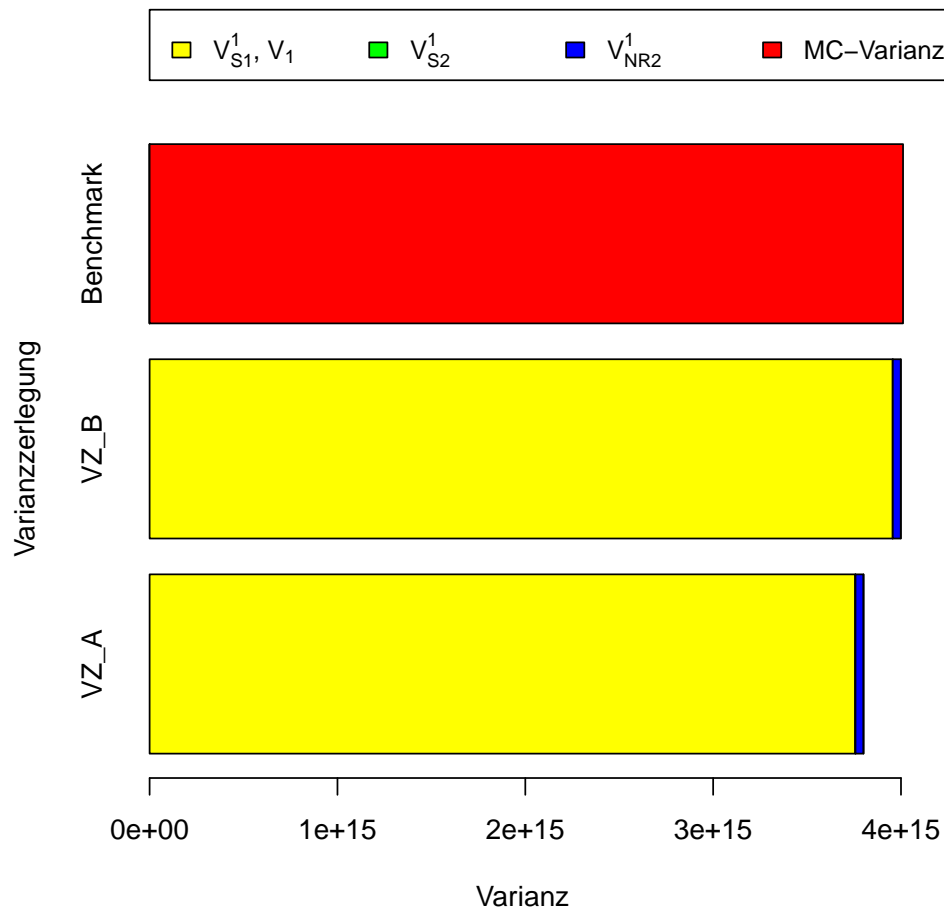


Abbildung 5.49: Varianzzerlegung für den Fall kleiner Auswahlsätze, heterogener PSU und einer ungleichen PSU-Größe

5.4.4.4 Varianzzerlegung für den Fall heterogener PSU, einer gleichen PSU-Größe und großer Auswahlsätze

Wie im Falle des einstufigen Stichprobendesigns gewinnt auch im mehrstufigen Fall die Varianzkomponente bezüglich fehlender Werte mit steigendem Gesamtauswahlsatz an Bedeutung. Daher besitzt die Komponente V_{NR2}^1 in Abbildung 5.50 einen nicht zu vernachlässigenden Anteil.⁴⁶ Da heterogene PSU und eine gleiche PSU-Größe vorliegen, ist die Komponente V_{S2}^1 größer als die Komponente V_{S1}^1 . Auch hier zeigt der Vergleich zwischen VZ_B und VZ_B den Einfluss der Imputation auf die Varianz der ersten Stufe.

⁴⁶Vgl. hierzu die Parameterkonstellationen des Szenarios in der Datensatzbeschreibung in Abschnitt A.2.2.22 im Anhang A.

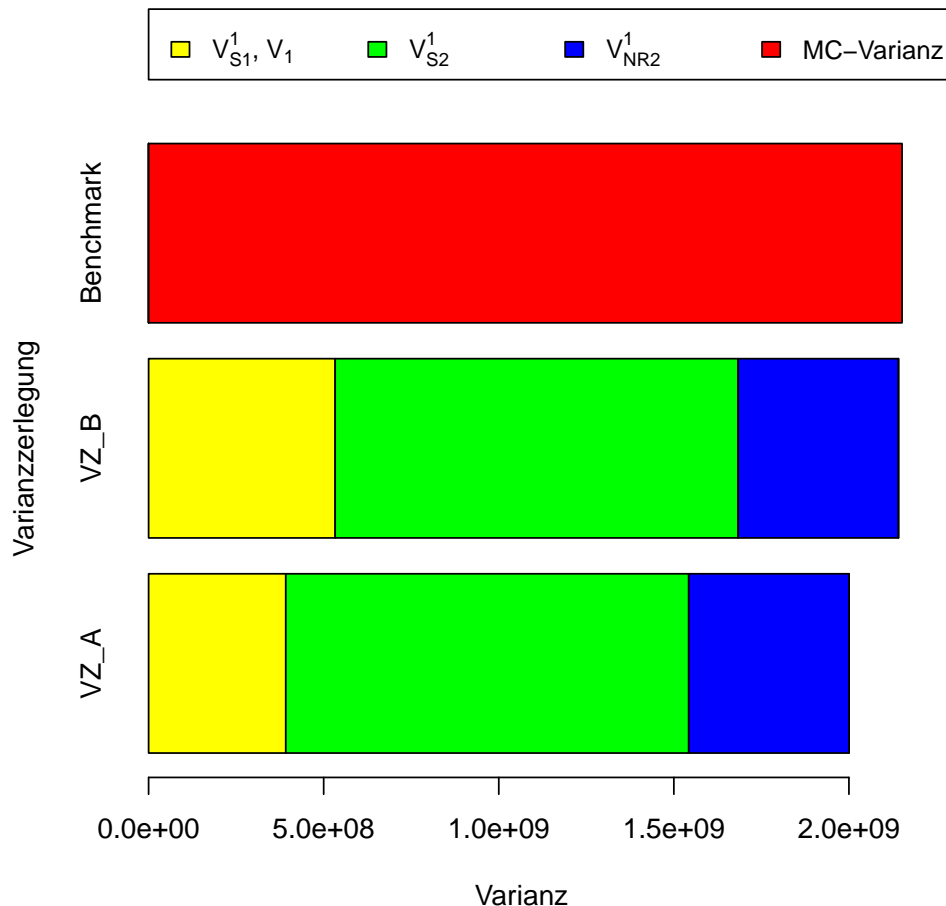


Abbildung 5.50: Varianzzerlegung für den Fall großer Auswahlsätze, heterogener PSU und einer ungleichen PSU-Größe

5.4.5 Fazit bezüglich der Behandlung mehrstufiger Stichprobendesigns

Im zweistufigen Stichprobendesign mit Stratifizierung zeigten sich die Vorteile des Rescaling-Bootstrap ohne Zurücklegen. Dies gilt insbesondere gegenüber dem BRR, welcher im Falle homogener PSU Schwierigkeiten offenbarte (vgl. Abschnitt 5.4.1.3). Der Grund liegt hier in der Gruppierung der homogenen PSU auf der ersten Stufe.

Über die Szenarien hinweg, zeigte der in dieser Arbeit in Abschnitt 4.5.5 vorgeschlagene Ansatz zur Übertragung des mehrstufigen Rescaling-Bootstrap ohne Zurücklegen auf den Fall fehlender Werte und Anwendung von Imputation nahezu unverzerrte Schätzungen. Jedoch muss hier beachtet werden, dass die Simulationen auf Basis eines normalverteilten Untersuchungsmerkmals durchgeführt worden sind. Die Ergebnisse des einstufigen Stichprobendesigns mit Stratifizierung in Abschnitt 5.2.4 zeigten die Schwierigkeiten der Anwendung von Varianzschätzmethoden im Falle von schiefen Verteilungen des Untersuchungsmerkmals. Die Differenzen bezüglich der Schätzungen basierend auf den in Abschnitt 4.5.5 vorgeschlagenen beiden Varianten der Kategorienbildung waren gering. Der Ansatz einer separaten Reimputation bzw. Adjustierung in den Kategorien K1, K3 und der zusammengelegten Kategorie K2 und K4 zeigte besonders im Falle von Hot-Deck-Imputation und großen Auswahlätzen exaktere Schätzungen. Dies könnte das Resultat der besseren Ausnutzung des Substichprobenbildungsprozesses sein. Die separate Reim-

putation bzw. Adjustierung zeichnete sich in manchen Szenarien durch weniger Ausreißer aus (vgl. z.B. Abschnitt 5.4.1.1). Hier könnte die ähnlichere Kategoriengröße von Vorteil sein. Eine eindeutige Handlungsempfehlung kann jedoch nicht gegeben werden. Hierzu waren die Unterschiede bezüglich der Schätzungen basierend auf beiden Varianten zu gering. Die Verbesserungen der Schätzungen gegenüber anderen Möglichkeiten der Kategorienbildung und gegenüber Schätzern, welche die Reimputation bzw. Adjustierung innerhalb bestimmter Kategorien vernachlässigten, waren in einigen Szenarien jedoch deutlich.

Die Verwendung der Korrekturfaktoren zur Berücksichtigung insbesondere der Varianzkomponente bezüglich fehlender Werte sind auch im Falle mehrstufiger Stichprobendesigns und einer gleichen PSU-Größe sinnvoll. Jedoch scheinen im Falle einer ungleichen PSU-Größe die Abweichungen zum Fall der einfachen Zufallsstichprobe zu groß. Hier zeigten sich große Überschätzungen der Varianz.

Die Varianzzerlegung im Falle mehrstufiger Stichprobendesigns zeigte insbesondere den Einfluss von Imputationen der letzten Stufe auf die Varianz der übergeordneten Stufe. Die Stichprobenkomponente \tilde{V}_S^1 liegt hier auf zwei Stufen vor (V_{S1}^1 und V_{S2}^1). Je nachdem, ob homogene oder heterogene PSU, eine gleiche oder ungleiche PSU-Größe vorliegen, ergibt sich eine unterschiedliche Varianzaufteilung zwischen den Stufen. Die Varianzkomponente bezüglich fehlender Werte \tilde{V}_{NR}^1 liegt in den verschiedenen Szenarien nur auf der zweiten Stufe (V_{NR2}^1) vor. Fehlende Werte wurden nur auf dieser Stufe erzeugt. Wie im einstufigen Design ist diese Komponente von dem Gesamtauswahlsatz abhängig.

6 Handlungsempfehlungen und Ausblick

6.1 Wichtige Erkenntnisse und Handlungsempfehlungen

Das Ziel der vorliegenden Arbeit war die Untersuchung der Varianzschätzung im Falle des Vorliegens fehlender Werte und ihrer Kompensation über Imputation. Hierzu wurden verschiedene Methoden verglichen. Diese können in die Gruppen direkte Varianzschätzer und Resampling-Methoden eingeteilt werden.

Die Anwendung direkter Varianzschätzer im Fall fehlender Werte und anschließender Imputation gestaltet sich bei komplexen Stichprobendesigns schwierig. Schon bei stratifizierten einstufigen Designs treten bei Verletzungen der Annahmen bezüglich des Ausfallmechanismus bei einigen direkten Schätzern größere Verzerrungen in Erscheinung. Dies zeigten auch die Ergebnisse der Simulationsstudie dieser Arbeit. Daher wurde im Rahmen dieser Arbeit die Anwendung von Resampling-Verfahren in den Fokus gestellt.

Unter den berücksichtigten Resampling-Verfahren zeigten insbesondere der Rescaling-Bootstrap ohne Zurücklegen und der BRR Vorteile. Auch der Delete-a-Group-Jackknife führte in den Szenarien in Abschnitt 5.2 häufig zu unverzerrten Schätzungen. Jedoch besaß insbesondere der Rescaling-Bootstrap ohne Zurücklegen Vorteile bezüglich der Effizienz gegenüber dem Delete-a-Group-Jackknife. Andere Resampling-Verfahren zeigten in den verschiedenen Szenarien Über- und Unterschätzungen der Varianz. Beim Shao/Sitter-Bootstrap, Repeated-Half-Sample-Bootstrap und bei der Random-Group-Methode sind Überschätzungen die Konsequenz einer fehlenden Endlichkeitskorrektur für den Fall des Ziehens ohne Zurücklegen. Beim unabhängigen Rescaling-Bootstrap sind Annahmeverletzungen und die Problematik der Wahl der Substichprobengröße ausschlaggebend bezüglich der auftretenden Verzerrungen.

Daher wurde im Rahmen dieser Arbeit die Anwendung des Rescaling-Bootstrap ohne Zurücklegen und des BRR bevorzugt. Diese Ansätze zeichnen sich in der in dieser Arbeit verwendeten Form dadurch aus, dass auch die nicht für die Substichprobe gezogenen Einheiten bei großen Auswahlätzen Einfluss auf die Berechnung des Punktschätzers der Resampling-Methode nehmen können. Dadurch beeinflussen sie die Varianzschätzung. In dieser Arbeit wurde Ansätze vorgeschlagen, wie auch solche Resampling-Verfahren auf den Fall fehlender Werte und deren Kompensation über Imputation übertragen werden können. Dies betrifft sowohl einstufige wie mehrstufige Stichprobendesigns. Die Ansätze basieren auf der Reimputation bzw. Adjustierung der Imputationswerte auch von nicht für die Substichprobe gezogener Elementen. Bei mehrstufigen Stichprobendesigns gilt es, zusätzlich die verschiedenen Stufen des Stichprobendesigns zu berücksichtigen. Ein wichtiger Punkt ist, dass die Imputationen untergeordneter Stufen Einfluss auf die Varianz übergeordneter Stufen besitzen.

Eine Handlungsempfehlung bezüglich des BRR und des Rescaling-Bootstrap ohne Zurücklegen kann aufgrund der durchgeführten Simulationsstudie in Abhängigkeit des Untersu-

chungsgegenstandes gegeben werden. Der BRR eignet sich besser für komplexe Imputationsmethoden wie Nearest-Neighbour-Imputation oder Predictive-Mean-Matching. Der Grund hierfür liegt in seiner ausbalancierten Auswahl der Substichprobe. Hierdurch ist die Methode weniger anfällig gegenüber Ausreißern. Dies kann insbesondere bei der Auswahl des nächsten Nachbarn in der Substichprobe von Vorteil sein. In der Arbeit wurde jedoch im Rahmen der Simulationsstudie ein Ansatz vorgeschlagen, mit dem auch der Rescaling-Bootstrap ohne Zurücklegen für Predictive-Mean-Matching angewendet werden kann. Dieser Ansatz sieht die Bestimmung der prognostizierten Werte des zugrunde liegenden Regressionsmodells vor der Substichprobenziehung vor.

Demgegenüber sollte der Rescaling-Bootstrap ohne Zurücklegen im Falle mehrstufiger Stichprobendesigns angewendet werden. Dieser besitzt hier gegenüber dem BRR den Vorteil, dass die Berücksichtigung der verschiedenen Stufen des Stichprobendesigns über die Gewichtung innerhalb der Berechnung des Rescaling-Punktschätzers erfolgt. Dadurch ist die Methode weniger aufwendig. Der BRR muss hingegen auf beiden Stufen separat angewendet werden. Der in dieser Arbeit für den BRR verwendete Ansatz geht zudem von einem konstanten Auswahlssatz an PSU über die Schichten aus, was das Einsatzgebiet weiter beschränkt. Zum anderen weist der BRR bei homogenen PSU größere Nachteile auf. Durch seine Gruppierung der PSU auf der ersten Stufe kann die Methode im Falle homogener PSU sehr variabel werden. Dies kann in bestimmten Fällen mit größeren Verzerrungen einhergehen, wie in Abschnitt 5.4.1.3 dargestellt.

Im Zusammenhang mit dem mehrstufigen Rescaling-Bootstrap wurde in der Arbeit ein Ansatz vorgeschlagen, welcher dieses Verfahren auf den Fall fehlender Werte und die Anwendung von Imputation überträgt. Dieser Ansatz basiert auf einer Bildung von Kategorien, innerhalb derer Reimputationen und Adjustierungen der imputierten Werte innerhalb der Substichproben durchzuführen sind. Diese Kategorienbildung soll zum einen eine Trennung für die Substichprobe gezogener und nicht gezogener Elemente beinhalten. Zum anderen sollen die Kategorien die verschiedenen Stufen des Stichprobendesigns berücksichtigen. In Abschnitt 4.5.5 wurden zwei Varianten der Kategorienbildung vorgeschlagen. Die Schätzer, welche auf den beiden Varianten basieren, zeigten in der Simulationsstudie keine großen Unterschiede. Es wurden in den Szenarien nahezu unverzerrte Schätzungen realisiert im Gegensatz zu Schätzern, welche auf anderen Varianten der Kategorienbildung basieren. Werden Reimputationen bzw. Adjustierungen innerhalb aller vier Kategorien durchgeführt, wurden in manchen Szenarien Schätzungen mit weniger Ausreißer realisiert (vgl. z.B. Abschnitt 5.4.1.1). Die ähnlicheren Kategoriengrößen sorgten hier für eine stabilere Varianzschätzung. Die Zusammenfassung der Kategorien K2 und K4 führt zu einer besseren Ausnutzung des Substichprobenziehungsprozesses des Rescaling-Bootstrap ohne Zurücklegen. Hier zeigt sich z.B. im Falle der Hot-Deck-Random-Imputation bei großen Auswahlssätzen in Abschnitt 5.4.2.2 eine exaktere Varianzschätzung. Aufgrund der marginalen Unterschiede kann jedoch zwischen den beiden Varianten keine Handlungsempfehlung gegeben werden.

Ein wichtiger Punkt dieser Arbeit war die Varianzzerlegung, insbesondere bezüglich der drei Fälle der Entstehung fehlender Werte. In diesem Zusammenhang wurde zusätzlich auch simulativ eine Varianzzerlegung durchgeführt. Hier wurde sowohl im einstufigen wie im mehrstufigen Stichprobendesign gezeigt, von welchen Parametern die einzelnen Komponenten abhängig sind. Diese Varianzzerlegungen sind im Bezug auf Resampling-Verfahren wichtig, die nur bestimmte Komponenten der Gesamtvarianz abdecken können. Nach MASHREGHI et al. (2014) sind die Methoden nicht in der Lage, die Varianzkomponente bezüglich fehlender Werte V_{NR}^1 bzw. \tilde{V}_{NR}^1 abzudecken und es zeigt sich im Falle großer Auswahlssätze eine schlechte Approximationsgüte der gemeinsamen Verteilung aus

Stichproben und Donoren.

Bezüglich der Erfassung dieser Varianzkomponenten werden in MASHREGHI et al. (2014) Korrekturfaktoren vorgeschlagen, welche multiplikativ mit der Resampling-Methode verknüpft werden. Diese müssen im Falle großer Auswahlätze berücksichtigt werden. MASHREGHI et al. (2014) gehen jedoch vom Fall variabler fehlender Werte in der Grundgesamtheit aus. Im Rahmen dieser Arbeit wurde zusätzlich ein Korrekturfaktor für den Fall eines fixen Response-Vektors in der Grundgesamtheit angegeben. Alle diese Korrekturfaktoren wurden jedoch unter den Bedingungen der einfachen Zufallsstichprobe hergeleitet. Die Handlungsempfehlung für die Anwendung der Korrekturfaktoren bezieht sich daher auf Stichprobendesigns bei denen Varianzen nicht zu sehr von der einfachen Zufallsstichprobe abweichen. Im Rahmen der Simulationsstudie wurde exemplarisch aufgezeigt, dass der Einbezug dieser Korrekturfaktoren im Falle mehrstufiger Stichprobendesigns mit einer ungleichen PSU-Größe mit Überschätzungen der Varianz einhergehen.

6.2 Ausblick

Die Resultate dieser Arbeit bieten Ansatzpunkte für weitergehende Forschung. Wie erläutert besitzt der BRR Vorteile bei komplexen Imputationsmethoden, der Rescaling-Bootstrap ohne Zurücklegen hingegen Vorteile bei komplexen Stichprobendesigns. Die Frage, die sich hier stellt, ist, welche Methode z.B. im Falle des Vorliegens eines mehrstufigen Stichprobendesigns und Nearest-Neighbour-Imputation verwendet werden kann. Hier wäre eine Kombination der beiden Resampling-Methoden denkbar. In diesem Fall wären detailliertere Untersuchungen notwendig, ob beispielsweise Substichproben für den Rescaling-Bootstrap ohne Zurücklegen ausbalanciert gezogen werden können. Ansätze hierfür bieten die Arbeiten zum Balanced-Bootstrap von DAVISON et al. (1986), GRAHAM et al. (1990) oder NIGAM und RAO (1996).

Auch die Erfassung der Varianz bezüglich fehlender Werte im Falle komplexer Stichprobendesigns muss weiter erforscht werden. Wenn die Abweichung zwischen den Varianzen des komplexen Stichprobendesigns und der einfachen Zufallsstichprobe zu groß werden, kann die Verwendung der beschriebenen Korrekturfaktoren problematisch werden. Für solche Fälle sind weitere Ansätze notwendig.

Im Rahmen dieser Arbeit wurde die Zerlegung der Gesamtvarianz mittels einer Monte-Carlo-Simulation durchgeführt. Hier ergab sich z.B. im mehrstufigen Stichprobendesign das Problem, die Varianz, welche im Zusammenhang mit einer ungleichen PSU-Größe steht, von den übrigen Komponenten zu trennen. Wie in der Arbeit am Beispiel des Szenarios mit einer ungleichen PSU-Größe gesehen, wäre eine solche Zerlegung für detaillierte Analysen wichtig.

Des Weiteren haben die Szenarien im Falle der Erzeugung fehlender Werte in schwach besetzten Kategorien gezeigt, dass Ungenauigkeiten bezüglich der Varianzschätzung resultieren können. Hier werden stabilere Varianzschätzer benötigt.

Ein letzter Ansatz für weitere Forschung stellt das Auftreten von fehlenden Werten und deren Kompensation über Imputation im Falle von *Unequal-Probability-Sampling* dar. Die Anwendung von Resampling-Verfahren für solche Stichprobendesigns ist meist sehr komplex. Im Artikel von MATEI und TILLÉ (2005) werden bezüglich dieser Thematik direkte Varianzschätzer angegeben, welche jedoch auf vollständigen Beobachtungen basieren. Daher werden auch für solche Stichprobendesigns Varianzschätzer benötigt, welche zusätzlich das Auftreten fehlender Werte und deren Kompensation über Imputation berücksichtigen.

A Anhang

A.1 Herleitung der Korrekturfaktoren aus Abschnitt 4.7.3

A.1.1 Herleitung des Korrekturfaktors für Mittelwertimputation bei der Entstehung variabler fehlender Werte in der Grundgesamtheit aus (4.49)

Der Korrekturfaktor der Mittelwertimputation aus MASHREGHI et al. (2014) kann wie folgt hergeleitet werden bei Verwendung von $\Delta_y = \sum_{i=1}^N (y_i - \bar{y}_U)^2$:

$$\begin{aligned}\alpha_{MeanImp} &= \frac{V_S^1 + V_{NR}^1}{V_S^1} \\ &= \frac{\frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \Delta_y + \frac{1-p_R}{p_R} \cdot \frac{1}{N^2} \cdot \Delta_y}{\frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \Delta_y} \\ &= \frac{\frac{1-f}{p_R \cdot f} + \frac{1-p_R}{p_R}}{\frac{1-f}{p_R \cdot f}} \\ &= \frac{1-f + (1-p_R) \cdot f}{1-f} \\ &= \frac{1-f \cdot p_R}{1-f}\end{aligned}$$

A.1.2 Herleitung des Korrekturfaktors für Hot-Deck-Random-Imputation bei Entstehung variabler fehlender Werte in der Grundgesamtheit aus (4.58)

Der Korrekturfaktor der (gewichteten) Hot-Deck-Random-Imputation aus MASHREGHI et al. (2014) im Falle der Entstehung variabler fehlender Werte in der Grundgesamtheit kann wie folgt hergeleitet werden bei Verwendung von $\Delta_y = \sum_{i=1}^N (y_i - \bar{y}_U)^2$:

$$\begin{aligned}
\alpha_{HotDeck} &= \frac{\tilde{V}_S^1 + \tilde{V}_{NR}^1 + \tilde{V}_{Imp}^1}{\tilde{V}_S^1 + (1-f) \cdot \tilde{V}_{Imp}^1} \\
&= \frac{\frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \Delta_y + \frac{1-p_R}{p_R} \cdot \frac{1}{N^2} \cdot \Delta_y + \frac{1-p_R}{n \cdot N} \cdot \Delta_y}{\frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \Delta_y + (1-f) \cdot \frac{1-p_R}{n \cdot N} \cdot \Delta_y} \\
&= \frac{\frac{1-f}{p_R \cdot f} + \frac{1-p_R}{p_R} \cdot + \frac{1-p_R}{n} \cdot N}{\frac{1-f}{p_R \cdot f} + (1-f) \cdot \frac{1-p_R}{n} \cdot N} \\
&= \frac{\frac{1-f}{p_R} + \frac{f \cdot (1-p_R)}{p_R} \cdot + 1 - p_R}{\frac{1-f}{p_R} + (1-f) \cdot (1-p_R)} \\
&= \frac{1-f + f \cdot (1-p_R) \cdot + (1-p_R) \cdot p_R}{1-f + (1-f) \cdot (1-p_R) \cdot p_R} \\
&= \frac{1 + p_R \cdot (1-p_R) - p_R \cdot f}{[1 + p_R \cdot (1-p_R)] \cdot (1-f)}
\end{aligned}$$

A.1.3 Herleitung des Korrekturfaktors für Hot-Deck-Random-Imputation für den Fall eines fixen Response-Vektors in der Grundgesamtheit aus (4.64)

Der Korrekturfaktor der (gewichteten) Hot-Deck-Random-Imputation aus dieser Arbeit im Falle eines fixen Response-Vektors in der Grundgesamtheit kann wie folgt hergeleitet werden bei Verwendung von $\Delta_y = \sum_{i=1}^N (y_i - \bar{y}_U)^2$:

$$\begin{aligned}
\alpha_{HotDeck} &= \frac{\tilde{V}_S^1 + \tilde{V}_{Imp}^1}{\tilde{V}_S^1 + (1-f) \cdot \tilde{V}_{Imp}^1} \\
&= \frac{\frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \Delta_y + \frac{1-p_R}{n \cdot N} \cdot \Delta_y}{\frac{1-f}{p_R \cdot f} \cdot \frac{1}{N^2} \cdot \Delta_y + (1-f) \cdot \frac{1-p_R}{n \cdot N} \cdot \Delta_y} \\
&= \frac{\frac{1-f}{p_R \cdot f} + \frac{1-p_R}{n} \cdot N}{\frac{1-f}{p_R \cdot f} + (1-f) \cdot \frac{1-p_R}{n} \cdot N} \\
&= \frac{\frac{1-f}{p_R} + 1 - p_R}{\frac{1-f}{p_R} + (1-f) \cdot (1-p_R)} \\
&= \frac{1-f + (1-p_R) \cdot p_R}{1-f + (1-f) \cdot (1-p_R) \cdot p_R} \\
&= \frac{1 + \frac{(1-p_R) \cdot p_R}{(1-f)}}{1 + (1-p_R) \cdot p_R}
\end{aligned}$$

A.2 Simulationsbeschreibung

In diesem Abschnitt wird eine kurze Übersicht und Beschreibung der in der Simulationsstudie verwendeten Methoden gegeben. Des Weiteren findet im Anschluss eine detaillierte Beschreibung der Szenarien statt.

A.2.1 Methodenübersicht

Die nachfolgenden Seiten beinhalten Tabellen, in welchen eine Beschreibung der in der Simulationsstudie berücksichtigten Methoden vorgenommen wird. Es erfolgt eine Trennung zwischen den Methoden, welche im einstufigen Design relevant sind und denjenigen welche im mehrstufigen Stichprobendesign von Interesse sind.

A.2.1.1 Einstufiges Stichprobendesign

Methode	Beschreibung
BRR	Methode des Balanced-Repeated-Replication, wie in Abschnitt 4.6.3 beschrieben. Reimputationen bzw. Adjustierungen erfolgen separat für die Imputationswerte zwischen für die Substichprobe gezogenen und nicht gezogenen Elementen
BRR_KF_zfix	Methode des Balanced-Repeated-Replication unter Berücksichtigung des Korrekturfaktors für den Fall eines fixen Response-Vektors in der Grundgesamtheit und Hot-Deck-Random-Imputation
BRR_KF	Methode des Balanced-Repeated-Replication unter Berücksichtigung des entsprechenden Korrekturfaktors für den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit
Dir	Direkter Varianzschätzer aus Abschnitt 3.6
Dir_Rao1	Direkter Varianzschätzer aus (3.47)
Dir_Rao2	Direkter Varianzschätzer aus (3.48)
Dir_Shao/Steel	Direkter Varianzschätzer aus (3.49)
Group	Delete-a-Group-Jackknife aus (4.17). Reimputationen bzw. Adjustierungen erfolgen in Imputationsklassen innerhalb der einzelnen Schichten
RG	Methode der Random-Groups, beschrieben in Abschnitt 4.2
Reg	Regressionsschätzer aus (3.52)
Repeated-Half	Repeated-Half-Sample-Bootstrap aus Abschnitt 4.4.2
Resc.oZ.M1	Rescaling-Bootstrap ohne Zurücklegen Modifikation 1, daher mit ausschließlicher Reimputation bzw. Adjustierung der Imputationswerte für die Substichprobe gezogener Elemente, beschrieben in 4.5.4
Resc.oZ.M1_KF_zfix	Rescaling-Bootstrap ohne Zurücklegen Modifikation 1 und Berücksichtigung des Korrekturfaktors für den Fall eines fixen Response-Vektors in der Grundgesamtheit und Hot-Deck-Random-Imputation

Methode	Beschreibung
Resc.oZ_M1_KF	Rescaling-Bootstrap ohne Zurücklegen Modifikation 1 und Berücksichtigung des entsprechenden Korrekturfaktors für den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit
Resc.oZ_M2	Rescaling-Bootstrap ohne Zurücklegen Modifikation 2, daher mit separater Reimputation bzw. Adjustierung der Imputationswerte zwischen für die Substichprobe gezogenem und nicht gezogenen Elementen, beschrieben in 4.5.4
Resc.oZ_M2_Adjust	Rescaling-Bootstrap ohne Zurücklegen Modifikation 2 unter Verwendung der Adjustierung der imputierten Werte
Resc.oZ_M2_KF_zfix	Rescaling-Bootstrap ohne Zurücklegen Modifikation 2 und Berücksichtigung des Korrekturfaktors für den Fall eines fixen Response-Vektors in der Grundgesamtheit und Hot-Deck-Random-Imputation
Resc.oZ_M2_KF	Rescaling-Bootstrap ohne Zurücklegen Modifikation 2 und Berücksichtigung des entsprechenden Korrekturfaktors für den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit
Resc.oZ_M2_PM	Rescaling-Bootstrap ohne Zurücklegen Modifikation 2 im Fall von Predictive-Mean-Matching und Berechnung der prognostizierten Werte vor der Substichprobenziehung, damit auf Basis der gesamten Stichprobe
Resc.oZ_M2_Reimp	Rescaling-Bootstrap ohne Zurücklegen Modifikation 2 unter Verwendung der Reimputation

Methode	Beschreibung
Resc_Unab1	Unabhängiger Rescaling-Bootstrap beschrieben in 4.7.2 mit einer Substichprobengröße in Höhe von $n_h - 3$.
Resc_Unab2	Unabhängiger Rescaling-Bootstrap beschrieben in 4.7.2 mit einer Substichprobengröße in Höhe von $\lfloor n_h/2 \rfloor$.
Resc_Unab3	Unabhängiger Rescaling-Bootstrap beschrieben in 4.7.2 mit einer Substichprobengröße aus (4.22).
Shao/Sitter	Shao und Sitter-Bootstrap, beschrieben in Abschnitt 4.4.1

A.2.1.2 Mehrstufiges Stichprobendesign

Methode	Beschreibung
BRR_1.2	Methode des Balanced-Repeated-Replication, wie in Abschnitt 4.6.3 beschrieben. Reimputationen bzw. Adjustierungen erfolgen separat für die Substichprobe gezogenen und nicht gezogenen Elementen, sowie auf beiden Stufen des Stichprobendesigns
BRR_1.2_KF	Methode des Balanced-Repeated-Replication. Reimputationen bzw. Adjustierungen erfolgen separat für die Imputationswerte zwischen für die Substichprobe gezogenen und nicht gezogenen Elementen, sowie auf beiden Stufen des Stichprobendesigns unter Berücksichtigung des entsprechenden Korrekturfaktors für den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit
BRR_2	Methode des Balanced-Repeated-Replication, wie in Abschnitt 4.6.3 beschrieben. Reimputationen bzw. Adjustierungen erfolgen separat für die Imputationswerte zwischen für die Substichprobe gezogenen und nicht gezogenen Elementen, jedoch nur auf der letzten Stufe des Stichprobendesigns
Resc_K1_K2_K3	Rescaling-Bootstrap ohne Zurücklegen mit ausschließlicher separater Reimputation bzw. Adjustierung in den Kategorien K1, K2 und K3. Keine Reimputation bzw. Adjustierung in K4
Resc_K1_K2_K3_KF	Rescaling-Bootstrap ohne Zurücklegen mit ausschließlicher separater Reimputation bzw. Adjustierung in den Kategorien K1, K2 und K3 unter Berücksichtigung des entsprechenden Korrekturfaktors für den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit. Keine Reimputation bzw. Adjustierung in K4
Resc_K1_K2_K3_K4	Rescaling-Bootstrap ohne Zurücklegen mit separater Reimputation bzw. Adjustierung in den Kategorien K1, K2, K3 und K4, beschrieben in 4.5.5

Methode	Beschreibung
Resc_K1_K2_K3_K4_KF	Rescaling-Bootstrap ohne Zurücklegen mit separater Reimputation bzw. Adjustierung in den Kategorien K1, K2, K3 und K4, unter Berücksichtigung des entsprechenden Korrekturfaktors für den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit
Resc_K1_K3	Rescaling-Bootstrap ohne Zurücklegen mit separater Reimputation bzw. Adjustierung ausschließlich in den Kategorien K1, K3. Keine Reimputation bzw. Adjustierung in K2 und K4
Resc_K1_K3_KF	Rescaling-Bootstrap ohne Zurücklegen mit separater Reimputation bzw. Adjustierung ausschließlich in den Kategorien K1, K3 unter Berücksichtigung des entsprechenden Korrekturfaktors für den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit. Keine Reimputation bzw. Adjustierung in K2 und K4
Resc_K1_K3_K2K4	Rescaling-Bootstrap ohne Zurücklegen mit separater Reimputation bzw. Adjustierung in den Kategorien K1, K3, sowie in den zusammengefassten Kategorien K2K4, beschrieben in 4.5.5
Resc_K1_K3_K2K4_KF	Rescaling-Bootstrap ohne Zurücklegen mit separater Reimputation bzw. Adjustierung in den Kategorien K1, K3, sowie in den zusammengefassten Kategorien K2K4 unter Berücksichtigung des entsprechenden Korrekturfaktors für den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit
Resc_K1K2K3_K4	Rescaling-Bootstrap ohne Zurücklegen mit separater Reimputation bzw. Adjustierung in den zusammengefassten Kategorie K1, K2 und K3 sowie K4

Methode	Beschreibung
Resc_K1K3_K2K4	Rescaling-Bootstrap ohne Zurücklegen mit separater Reimputation bzw. Adjustierung in den zusammengefassten Kategorien K1 und K3 sowie K2 und K4
Resc_K2_K3_KF	Rescaling-Bootstrap ohne Zurücklegen mit separater Reimputation bzw. Adjustierung ausschließlich in den Kategorien K2 und K3 unter Berücksichtigung des entsprechenden Korrekturfaktors für den Fall der Entstehung variabler fehlender Werte in der Grundgesamtheit. Keine Reimputation bzw. Adjustierung in K1 und K4
Resc_K2K3_K1K4	Rescaling-Bootstrap ohne Zurücklegen mit separater Reimputation bzw. Adjustierung in den zusammengefassten Kategorien K2 und K3 sowie K1 und K4

A.2.2 Szenarien

A.2.2.1 Parameterkonstellationen zum Vergleich der Varianzen der imputierten Werte nach den verschiedenen Imputationsmethoden aus Abschnitt 2.2.9

- *Grundgesamtheit*: Region 3 aus AMELIA in Abschnitt 5.1.1 beschrieben (821.472 Haushalte).
- *Untersuchungsmerkmal*: Einkommen
- *Hilfsvariable*: 1. Grundgesamtheit: Metrische Hilfsvariable aus Modell (5.1) mit den Parametern $\alpha = 100$, $\beta = 5$, $\delta = 2.000.000$, $\mu = 12.000.000$ und $\sigma_u = 2.200.000$ und einer Korrelation zum Untersuchungsmerkmal in Höhe von 11 %. 2. Grundgesamtheit: Metrische Hilfsvariable aus Modell (5.1) mit den Parametern $\alpha = 100$, $\beta = 5$, $\delta = 2.000.000$, $\mu = 12.000.000$ und $\sigma_u = 25.000$ und einer Korrelation zum Untersuchungsmerkmal in Höhe von 70 %.
- *Stichprobendesign*: Geschichtete Zufallsstichprobe (vgl. Abschnitt 5.1.1)
- *Stichprobenumfang*: Insgesamt bei 40.000 Haushalten, Gesamtauswahlsatz bei knapp 5 %
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: 1. Grundgesamtheit: In Abhängigkeit der nach der nach Modell (5.1) erzeugten metrischen Hilfsvariable (MAR) nach Formel (5.5) mit $\tilde{\alpha} = 15$ und $\tilde{\beta} = -2.393$. 2. Grundgesamtheit: In Abhängigkeit der nach Modell (5.1) erzeugten metrischen Hilfsvariable (MAR) nach Formel (5.5) mit $\tilde{\alpha} = 15$ und $\tilde{\beta} = -2.42$.
- *Ausfallrate*: insgesamt bei ca. 46 %
- *Imputationsklassen*: Imputationen werden für alle Imputationsmethoden aus Gründen der Vergleichbarkeit innerhalb der Schichten durchgeführt.

A.2.2.2 Mittelwertimputation Abschnitt 5.2.1.1

- *Grundgesamtheit*: Region 3 aus AMELIA in Abschnitt 5.1.1 beschrieben (821.472 Haushalte).
- *Untersuchungsmerkmal*: Einkommen
- *Hilfsvariable*: Haushaltstyp
- *Stichprobendesign*: Geschichtete Zufallsstichprobe (vgl. Abschnitt 5.1.1)
- *Stichprobenumfang*: Insgesamt bei 40.000 Haushalten, Gesamtauswahlsatz bei knapp 5 %
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit

- *Erzeugung fehlender Werte*: In Abhängigkeit des Haushaltstyps (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	-7.7	-10.1	-10.3	-10.0

- *Ausfallrate*: insgesamt bei ca. 21 %
- *Imputation*: Bildung der Imputationsklassen über den Haushaltstyp
- *Besonderheiten*: Für die beiden Szenarien am Ende des Abschnitts gelten die gleichen Parameterkonstellationen nur mit Bezug auf die Erzeugung fehlender Werte in der Stichprobe bzw. eines fixen Response-Vektors.

A.2.2.3 Verhältnisimputation aus Abschnitt 5.2.1.2

- *Grundgesamtheit*: Region 3 aus AMELIA in Abschnitt 5.1.1 beschrieben (821.472 Haushalte).
- *Untersuchungsmerkmal*: Einkommen
- *Hilfsvariable*: Metrische Hilfsvariable aus Modell (5.1)
- *Stichprobendesign*: Geschichtete Zufallsstichprobe (vgl. Abschnitt 5.1.1)
- *Stichprobenumfang*: Insgesamt bei 40.000 Haushalten, Gesamtauswahlsatz bei knapp 5 %
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der nach Modell (5.1) erzeugten metrischen Hilfsvariable (MAR) nach Formel (5.5) mit $\tilde{\alpha} = 15$ und $\tilde{\beta} = -1,04$
- *Ausfallrate*: insgesamt bei ca. 24 %
- *Imputationsklassen*: Keine Bildung von Imputationsklassen

A.2.2.4 Nearest-Neighbour-Imputation Abschnitt 5.2.2.2

Hier wurden zwei Szenarien verwendet.

1.Szenario der Nearest-Neighbour-Imputation:

- *Grundgesamtheit*: Region 3 aus AMELIA in Abschnitt 5.1.1 beschrieben (821.472 Haushalte).
- *Untersuchungsmerkmal*: Einkommen
- *Hilfsvariable*: Metrische Hilfsvariable aus Modell (5.1)
- *Stichprobendesign*: Geschichtete Zufallsstichprobe (vgl. Abschnitt 5.1.1)
- *Stichprobenumfang*: Insgesamt bei 40.000 Haushalten, Gesamtauswahlsatz bei knapp 5 %

- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der nach der nach Modell (5.1) erzeugten metrischen Hilfsvariable (MAR) nach Formel (5.5) mit $\tilde{\alpha} = 15$ und $\tilde{\beta} = -1,04$
- *Ausfallrate*: insgesamt bei ca. 24 %
- *Imputationsklassen*: Hier wird eine synthetisch erzeugte Variable verwendet, welche mit der metrischen Hilfsvariable aus (5.1) hoch korreliert ist.

2.Szenario der Nearest-Neighbour-Imputation:

- *Grundgesamtheit*: Region 3 aus AMELIA in Abschnitt 5.1.1 beschrieben (821.472 Haushalte).
- *Untersuchungsmerkmal*: Gesamtanzahl der Kategorie 2 der Variable Haushaltsgröße
- *Hilfsvariable*: Einkommen
- *Stichprobendesign*: Geschichtete Zufallsstichprobe (vgl. Abschnitt 5.1.1)
- *Stichprobenumfang*: Insgesamt bei 40.000 Haushalten, Gesamtauswahlsatz bei knapp 5 %
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der nach Modell (5.1) erzeugten metrischen Hilfsvariable (MAR) nach Formel (5.5) mit $\tilde{\alpha} = 3$ und $\tilde{\beta} = -0,00003$
- *Ausfallrate*: insgesamt bei ca. 24 %
- *Imputationsklassen*: Hier wird eine synthetisch erzeugte Variable verwendet, welche mit dem Einkommen aus (5.1) hoch korreliert ist.

A.2.2.5 Predictive-Mean-Matching Abschnitt 5.2.2.3

- *Grundgesamtheit*: Region 3 aus AMELIA in Abschnitt 5.1.1 beschrieben (821.472 Haushalte).
- *Untersuchungsmerkmal*: Einkommen
- *Hilfsvariablen*: Haushaltstyp und metrische Hilfsvariable aus Modell (5.1)
- *Stichprobendesign*: Geschichtete Zufallsstichprobe (vgl. Abschnitt 5.1.1)
- *Stichprobenumfang*: Insgesamt bei 40.000 Haushalten, Gesamtauswahlsatz bei knapp 5 %
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der nach Modell (5.1) erzeugten metrischen Hilfsvariable sowie dem Haushaltstyp (MAR) nach Formel (5.7) mit $\tilde{\alpha} = 4,5$, $\tilde{\beta} = 0,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	-7.7	-10.1	-10.3	-10.0

- *Ausfallrate*: insgesamt bei ca. 21 %
- *Imputationsklassen*: Imputationen werden innerhalb der Schichten durchgeführt

A.2.2.6 Erhöhung der Rate der fehlenden Werte Abschnitt 5.2.3.1

- *Grundgesamtheit*: Region 3 aus AMELIA in Abschnitt 5.1.1 beschrieben (821.472 Haushalte).
- *Untersuchungsmerkmal*: Einkommen
- *Hilfsvariable*: Haushaltstyp
- *Stichprobendesign*: Geschichtete Zufallsstichprobe (vgl. Abschnitt 5.1.1)
- *Stichprobenumfang*: Insgesamt bei 40.000 Haushalten, Gesamtauswahlsatz bei knapp 5 %
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit des Haushaltstyps (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 8,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	-7.7	-10.1	-10.3	-10.0

- *Ausfallrate*: insgesamt bei ca. 46 %
- *Imputation*: Bildung der Imputationsklassen über den Haushaltstyp. Mittelwertimputation

A.2.2.7 Erzeugung fehlender Werte in den gering besetzten Kategorien Abschnitt 5.2.3.2

- *Grundgesamtheit*: Region 3 aus AMELIA in Abschnitt 5.1.1 beschrieben (821.472 Haushalte).
- *Untersuchungsmerkmal*: Einkommen
- *Hilfsvariable*: Haushaltstyp
- *Stichprobendesign*: Geschichtete Zufallsstichprobe (vgl. Abschnitt 5.1.1)
- *Stichprobenumfang*: Insgesamt bei 40.000 Haushalten, Gesamtauswahlsatz bei knapp 5 %
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit des Haushaltstyps (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 15,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	-17	-17	2	2	1

- *Ausfallrate*: insgesamt bei ca. 20 %
- *Imputation*: Bildung der Imputationsklassen über den Haushaltstyp. Mittelwertimputation

A.2.2.8 Vergleich unterschiedlicher Ausfallmechanismen aus 5.2.3.3

Hier wurden die drei Szenarien MAR, MCAR und NMAR verglichen. Diese unterscheiden sich lediglich bezüglich der Generierung fehlender Werte. Für die übrigen Parameterkonstellationen siehe Abschnitt A.2.2.2.

Szenario MAR :

- *Erzeugung fehlender Werte*: siehe Abschnitt A.2.2.2
- *Ausfallrate*: insgesamt bei ca. 21 %

Szenario MCAR :

- *Erzeugung fehlender Werte*: Fehlende Werte werden zufällig aus der Grundgesamtheit gezogen ohne die Berücksichtigung von Antwortwahrscheinlichkeiten
- *Ausfallrate*: insgesamt bei ca. 21 %

Szenario NMAR :

- *Erzeugung fehlender Werte*: In Abhängigkeit nach dem Untersuchungsmerkmal Einkommen nach Formel (5.8) mit $\tilde{\alpha} = 10.5$ und $\tilde{\beta} = -0.000125$
- *Ausfallrate*: insgesamt bei ca. 21 %

A.2.2.9 Mittelwertimputation große Auswahlätze Abschnitt 5.2.4.1

- *Grundgesamtheit*: Region 3 aus AMELIA in Abschnitt 5.1.1 beschrieben (821.472 Haushalte).
- *Untersuchungsmerkmal*: Einkommen
- *Hilfsvariable*: Haushaltstyp
- *Stichprobendesign*: Geschichtete Zufallsstichprobe (vgl. Abschnitt 5.1.1)
- *Stichprobenumfang*: Insgesamt bei 380.000 Haushalten, Gesamtauswahlsatz bei ungefähr 46 %
- *Fall fehlender Werte*: Es werden hier drei Szenarien gebildet, welche die Fälle der Entstehung variabler fehlender Werte in der Grundgesamtheit, Entstehung fehlender Werte in der Stichprobe und eines fixen Response-Vektors berücksichtigen
- *Erzeugung fehlender Werte*: In Abhängigkeit des Haushaltstyps (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	-7.7	-10.1	-10.3	-10.0

- *Ausfallrate*: insgesamt bei ca. 21 %
- *Imputation*: Bildung der Imputationsklassen über den Haushaltstyp

A.2.2.10 Varianzzerlegung einfache Zufallsstichprobe 5.3

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Einfache Zufallsstichprobe
- *Stichprobenumfang*: Der Gesamtstichprobenumfang beträgt im Falle eines kleinen Auswahlsatzes 40.000 (Gesamtauswahlsatz von 1,6 %), bei einem mittlerem Auswahlsatz 125.000 (Gesamtauswahlsatz von 50 %) und bei einem großen Auswahlsatz 2.250.000 (Gesamtauswahlsatz von 90 %)
- *Fall fehlender Werte*: Abhängig vom betrachteten Szenario
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6). Im Falle einer kleinen Ausfallrate gilt $\tilde{\alpha} = 7.7$, bei mittlerer Ausfallrate beträgt $\tilde{\alpha} = 5.7$ im Falle einer großen Ausfallrate gilt $\tilde{\alpha} = 3.7$. Für $\tilde{\gamma}_i$ gilt in allen Szenarios:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: Im Falle der kleinen Ausfallrate bei ungefähr 5 %, bei der mittleren Ausfallrate bei ungefähr 22.5 % und bei der hohen Ausfallrate bei ungefähr 43 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable

A.2.2.11 Ergebnisse mehrstufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe für den Fall heterogener PSU, Mittelwertimputation, einer gleichen PSU-Größe und kleiner Auswahlsätze Abschnitt 5.4.1.1

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe. Gleiche Schicht- und PSU-Größe (50 PSU pro Schicht, 2.000 USU pro PSU). Heterogene PSU.
- *Stichprobenumfang*: $l_h = 5$ PSU pro Schicht auf der ersten Stufe und ein Gesamtstichprobenumfang von 25.000 Einheiten auf der zweiten Stufe

- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable

A.2.2.12 Ergebnisse mehrstufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe für den Fall von heterogener PSU, Mittelwertimputation, einer gleichen PSU-Größe und mittlerer Auswahlsätze Abschnitt 5.4.1.1

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe. Gleiche Schicht- und PSU-Größe (50 PSU pro Schicht, 2.000 USU pro PSU). Heterogene PSU.
- *Stichprobenumfang*: $l_h = 10$ PSU pro Schicht auf der ersten Stufe und ein Gesamtstichprobenumfang von 125.000 Einheiten auf der zweiten Stufe
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable

A.2.2.13 Ergebnisse mehrstufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe für den Fall homogener PSU, Mittelwertimputation, einer gleichen PSU-Größe und kleiner Auswahlsätze Abschnitt 5.4.1.3

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Nach der Größe sortiertes normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe. Gleiche Schicht- und PSU-Größe (50 PSU pro Schicht, 2.000 USU pro PSU). Homogene PSU.
- *Stichprobenumfang*: $l_h = 5$ PSU pro Schicht auf der ersten Stufe und ein Gesamtstichprobenumfang von 25.000 Einheiten auf der zweiten Stufe
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable

A.2.2.14 Ergebnisse mehrstufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe für den Fall heterogener PSU, Mittelwertimputation, einer gleichen PSU-Größe und großer Auswahlsätze Abschnitt 5.4.2.1

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe. Gleiche Schicht- und PSU-Größe (50 PSU pro Schicht, 2.000 USU pro PSU). Heterogene PSU.
- *Stichprobenumfang*: $l_h = 38$ PSU pro Schicht auf der ersten Stufe und ein Gesamtstichprobenumfang von 625.000 Einheiten auf der zweiten Stufe

- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable

A.2.2.15 Ergebnisse mehrstufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe für den Fall homogener PSU, Mittelwertimputation, einer gleichen PSU-Größe und großer Auswahlätze Abschnitt 5.4.2.3

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Sortiertes normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe. Gleiche Schicht- und PSU-Größe (50 PSU pro Schicht, 2.000 USU pro PSU). Homogene PSU.
- *Stichprobenumfang*: $l_h = 38$ PSU pro Schicht auf der ersten Stufe und ein Gesamtstichprobenumfang von 625.000 Einheiten auf der zweiten Stufe
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable

A.2.2.16 Ergebnisse mehrstufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe für den Fall heterogener PSU, Mittelwertimputation, einer gering variierenden PSU-Größe und kleiner Auswahlsätze Abschnitt 5.4.3.1

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe. Gering variierende Schicht- und PSU-Größe (Die Schichtgröße variiert von 46 bis 52 PSU pro Schicht, Die PSU-Größe variiert von 1.900 bis 2.100 USU pro PSU). Heterogene PSU.
- *Stichprobenumfang*: $l_h = 5$ PSU pro Schicht auf der ersten Stufe und ein Gesamtstichprobenumfang von 25.000 Einheiten auf der zweiten Stufe
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable

A.2.2.17 Ergebnisse mehrstufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe für den Fall heterogener PSU, Mittelwertimputation, einer stark variierende PSU-Größe und kleiner Auswahlsätze Abschnitt 5.4.3.1

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe. Stark variierende Schicht- und PSU-Größe (Die Schichtgröße variiert von 40 bis 60 PSU pro Schicht, die PSU-Größe variiert von 1.000 bis 3.000 USU pro PSU). Heterogene PSU.

- *Stichprobenumfang*: $l_h = 5$ PSU pro Schicht auf der ersten Stufe und ein Gesamtstichprobenumfang von 25.000 Einheiten auf der zweiten Stufe
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable

A.2.2.18 Ergebnisse mehrstufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe für den Fall heterogener PSU, Mittelwertimputation, einer stark variierende PSU-Größe und großer Auswahlätze Abschnitt 5.4.3.2

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufiges Stichprobendesign mit einer Schichtung auf der ersten Stufe. stark variierende Schicht- und PSU-Größe (Die Schichtgröße variiert von 40 bis 60 PSU pro Schicht, die PSU-Größe variiert von 1.000 bis 3.000 USU pro PSU). Heterogene PSU.
- *Stichprobenumfang*: $l_h = 38$ PSU pro Schicht auf der ersten Stufe und ein Gesamtstichprobenumfang von 625.000 Einheiten auf der zweiten Stufe
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable

A.2.2.19 Varianzzerlegung mehrstufiges Stichprobendesign ohne Schichtung für den Fall heterogener PSU, einer gleichen PSU-Größe und kleiner Auswahlsätze Abschnitt 5.4.4.1

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufige Zufallsstichprobe ohne Schichtung. Gleiche PSU-Größe (2.000 USU). Heterogene PSU.
- *Stichprobenumfang*: $l = 125$ PSU auf der ersten Stufe und ein Gesamtstichprobenumfang von 25.000 Einheiten auf der zweiten Stufe
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable. Anwendung von Mittelwertimputation

A.2.2.20 Varianzzerlegung mehrstufiges Stichprobendesign ohne Schichtung für den Fall homogener PSU, einer gleichen PSU-Größe und kleiner Auswahlsätze Abschnitt 5.4.4.2

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Sortiertes normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufige Zufallsstichprobe ohne Schichtung. Gleiche PSU-Größe (2.000 USU). Homogene PSU.
- *Stichprobenumfang*: $l = 125$ PSU auf der ersten Stufe und ein Gesamtstichprobenumfang von 25.000 Einheiten auf der zweiten Stufe
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit

- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable. Anwendung von Mittelwertimputation

A.2.2.21 Varianzzerlegung mehrstufiges Stichprobendesign ohne Schichtung für den Fall heterogener PSU, einer ungleichen PSU-Größe und kleiner Auswahlsätze Abschnitt 5.4.4.3

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufige Zufallsstichprobe ohne Schichtung. Ungleiche PSU-Größe (Die Größe variiert zwischen 1.000 USU und 3.000 USU). Heterogene PSU.
- *Stichprobenumfang*: $l = 125$ PSU auf der ersten Stufe und ein Gesamtstichprobenumfang von 25.000 Einheiten auf der zweiten Stufe
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable. Anwendung von Mittelwertimputation

A.2.2.22 Varianzzerlegung mehrstufiges Stichprobendesign ohne Schichtung für den Fall heterogener PSU, einer gleichen PSU-Größe und großer Auswahlätze Abschnitt 5.4.4.4

- *Grundgesamtheit*: Nach Abschnitt 5.1.2 synthetisch erzeugte Grundgesamtheit (2.500.000 Einheiten).
- *Untersuchungsmerkmal*: Normalverteiltes Untersuchungsmerkmal aus Abschnitt 5.1.2
- *Hilfsvariable*: Synthetisch erzeugte kategoriale Variable aus Abschnitt 5.1.2
- *Stichprobendesign*: Zweistufige Zufallsstichprobe ohne Schichtung. Gleiche PSU-Größe (2.000 USU). Heterogene PSU.
- *Stichprobenumfang*: $l = 950$ PSU auf der ersten Stufe und ein Gesamtstichprobenumfang von 1.250.000 Einheiten auf der zweiten Stufe
- *Fall fehlender Werte*: Entstehung variabler fehlender Werte in der Grundgesamtheit
- *Erzeugung fehlender Werte*: In Abhängigkeit der synthetisch erzeugten kategorialen Variable aus Abschnitt 5.1.2 (MAR) nach Formel (5.6) mit $\tilde{\alpha} = 10,5$ und $\tilde{\gamma}_i$ in Abhängigkeit der Kategorie:

Kategorie	1	2	3	4	5	6
$\tilde{\gamma}_i$	0.0	2.0	2.0	-5.5	-5.5	-5.5

- *Ausfallrate*: ungefähr 22.5 %.
- *Imputation*: Bildung der Imputationsklassen nach der synthetisch erzeugten kategorialen Variable. Anwendung von Mittelwertimputation

A.2.3 Programm zur Erzeugung der Schichten und PSU in R für das mehrstufige Stichprobendesign mit Stratifizierung auf der ersten Stufe

Das folgende R-Programm bezieht sich auf die Generierung der PSU und der Schichten im Falle einer stark variierenden PSU-Größe im mehrstufigen Stichprobendesign mit einer Schichtung auf der ersten Stufe. Es orientiert sich an der Erzeugung der PSU in BRUCH et al. (2011).

```
psu_size<-c(rep(2000,550),rep(1000,200),rep(3000,200),rep(1750,50),rep(2250,50),rep(1700,50),rep(2300,50),rep(2500,50),rep(1500,50))
psu<-rep(1:1250,sample(psu_size,size=length(psu_size),replace=F))
```

```
sch<-rep(NA,length(psu))
sch_size<-c(rep(40,5),rep(50,7),rep(60,5),rep(45,1),rep(55,1),rep(47,1),rep(53,1),rep(48,1),rep(52,1),rep(42,1),rep(58,1))
anzschGG<-25

schgrbez<-0

for (i in 1:anzschGG){

  schbez1<-c((schgrbez+1):sum(sch_size[1:i]))

  schgrbez<-sum(sch_size[1:i])

  sch[psu %in% schbez1]<-i

}
```

A.3 Simulationsergebnisse

A.3.1 Punktschätzer bezüglich des Vergleichs unterschiedlicher Ausfallmechanismen in Abschnitt 5.2.3.3

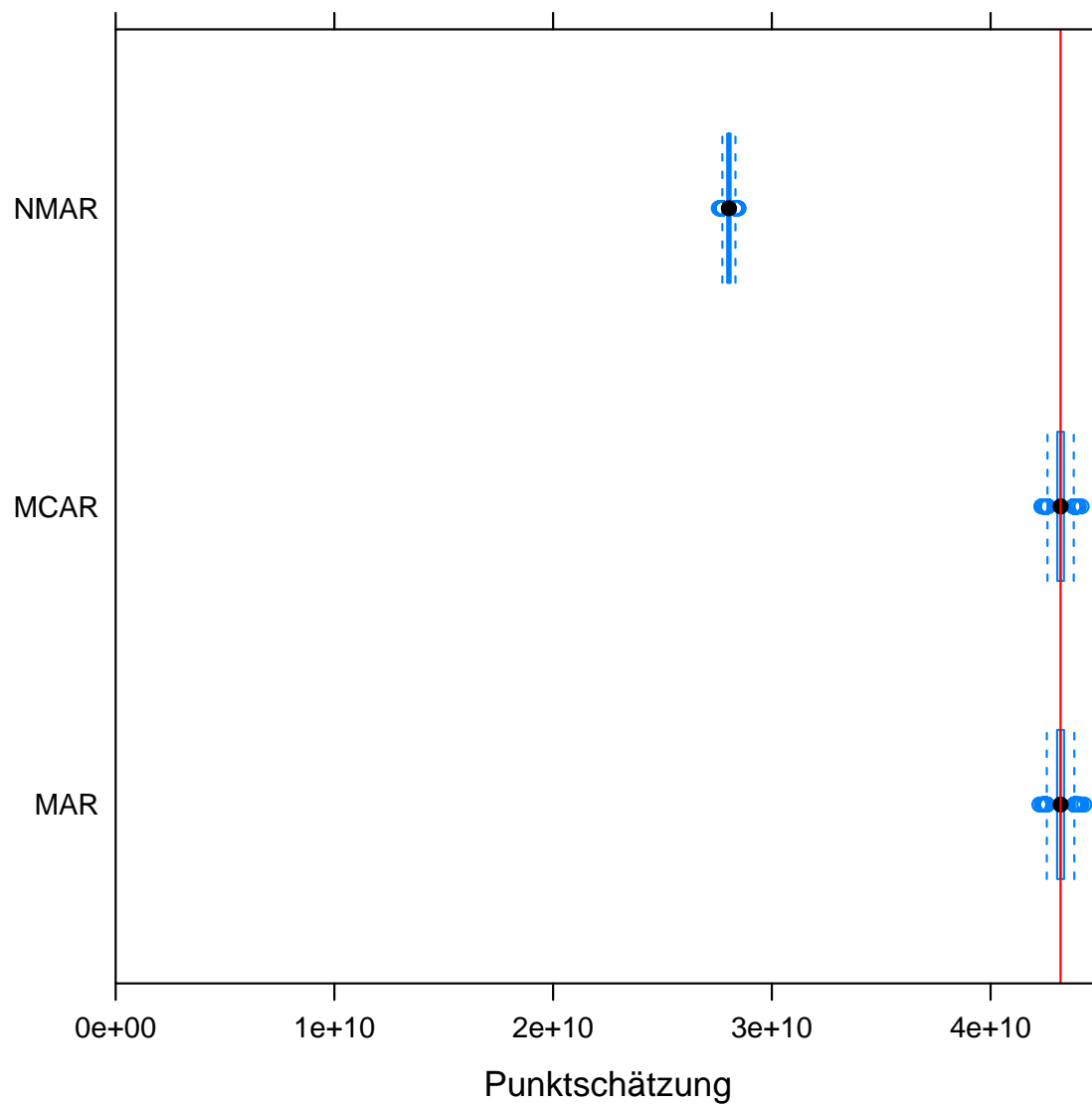


Abbildung A.1: Vergleich der Punktschätzer unter unterschiedlichen Ausfallmechanismen

Abbildung A.1 zeigt die Punktschätzungen unter den verschiedenen Ausfallmechanismen aus Abschnitt 5.2.3.3. Wie an der Abbildung zu sehen ist, ist der Punktschätzer im Falle von MCAR und MAR unverzerrt. Bei Vorliegen von NMAR ergeben sich große Unterschätzungen.

A.3.2 Vergleich der direkten Varianzschätzer im Falle von Regressionsimputation

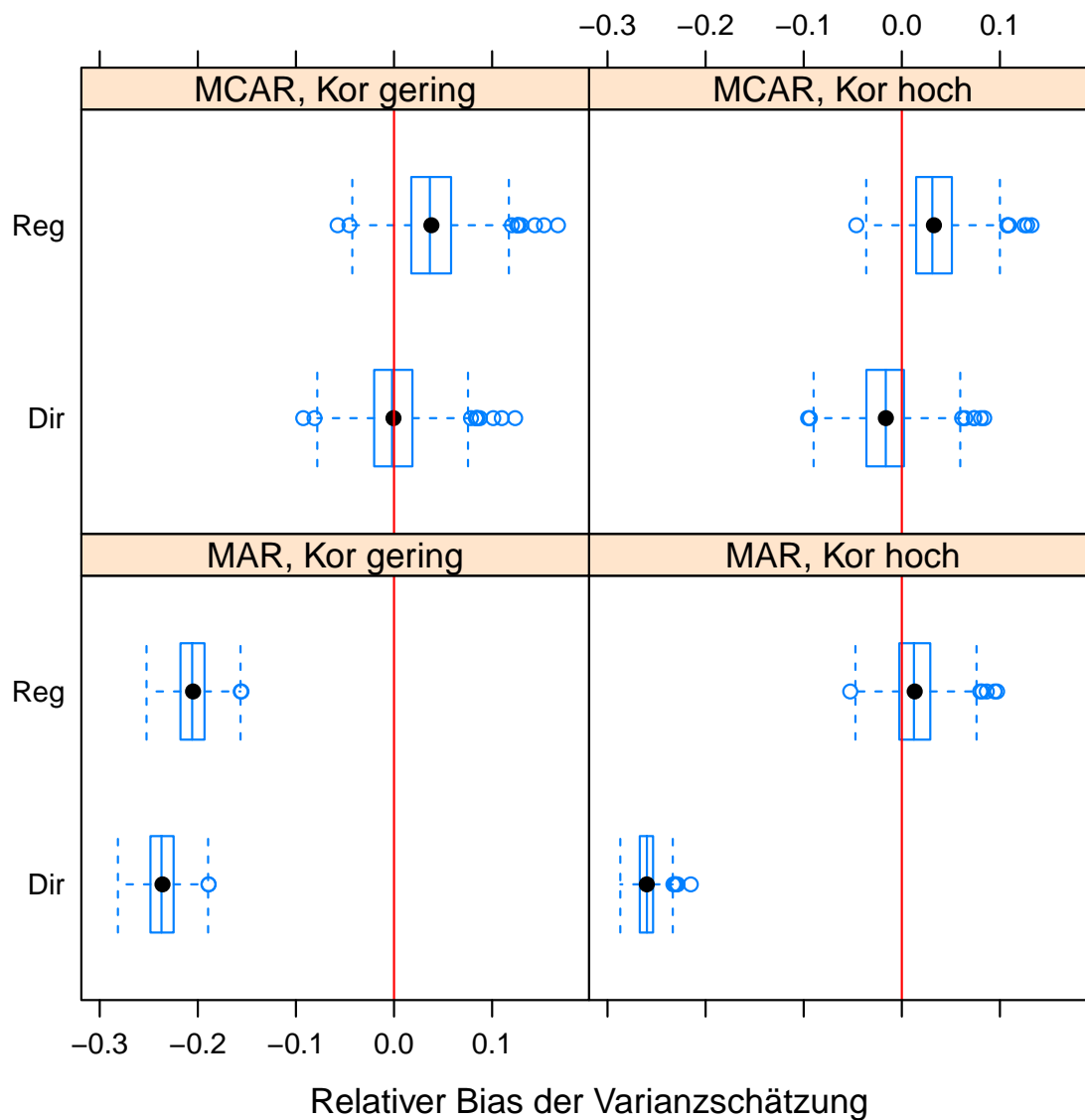


Abbildung A.2: Vergleich der direkten Varianzschätzer im Falle von Regressionsimputation und der einfachen Zufallsstichprobe

Abbildung A.2 zeigt die Anwendbarkeit der direkten Varianzschätzer für Regressionsimputation aus Abschnitt 3.6.4 im Falle der einfachen Zufallsstichprobe.¹ Dabei wird unterschieden zwischen den Ausfallmechanismen MAR und MCAR, sowie einer hohen und niedrigen Korrelation bezüglich x und y . Es werden der Schätzer aus (3.51), bezeichnet

¹ Es liegt grundsätzlich die Grundgesamtheit AMELIA aus Abschnitt 5.1.1 zugrunde mit dem Untersuchungsmerkmal Einkommen. Jedoch wird keine Schichtung vorgenommen. Der Stichprobenumfang beträgt 40.000 Haushalte. Der Ausfall wird nach einer metrischen Hilfsvariable erzeugt, welche nach Modell (5.1) in Abhängigkeit der Korrelation von y und x erzeugt wird. Im Falle der hohen Korrelation liegt die Korrelation bei ungefähr 95 %, bei einer niedrigen Korrelation bei ungefähr 10 %. Die Ausfallrate liegt 24 %. Nur im Szenario mit hoher Korrelation und MAR liegt sie bei 19 %.

durch *Dir* und der Schätzer aus (3.52) bezeichnet durch *Reg* verglichen. Letzterer stellt die Anwendung des Regressionsschätzers nach KIM (2001) dar. Die Ergebnisse zeigen, dass das Verfahren nur bei einem starken Regressionszusammenhang mit hoher Korrelation und im Falle von MAR verwendet werden kann. Der direkte Varianzschätzer aus (3.51) kann nur im Falle von MCAR benutzt werden. Hier liegt auch der uniforme Antwortmechanismus vor, welches die zentrale Annahme dieses Schätzers ist.

A.3.3 Regressionsimputation mehrstufiges Stichprobendesign mit einer Stratifizierung auf der ersten Stufe

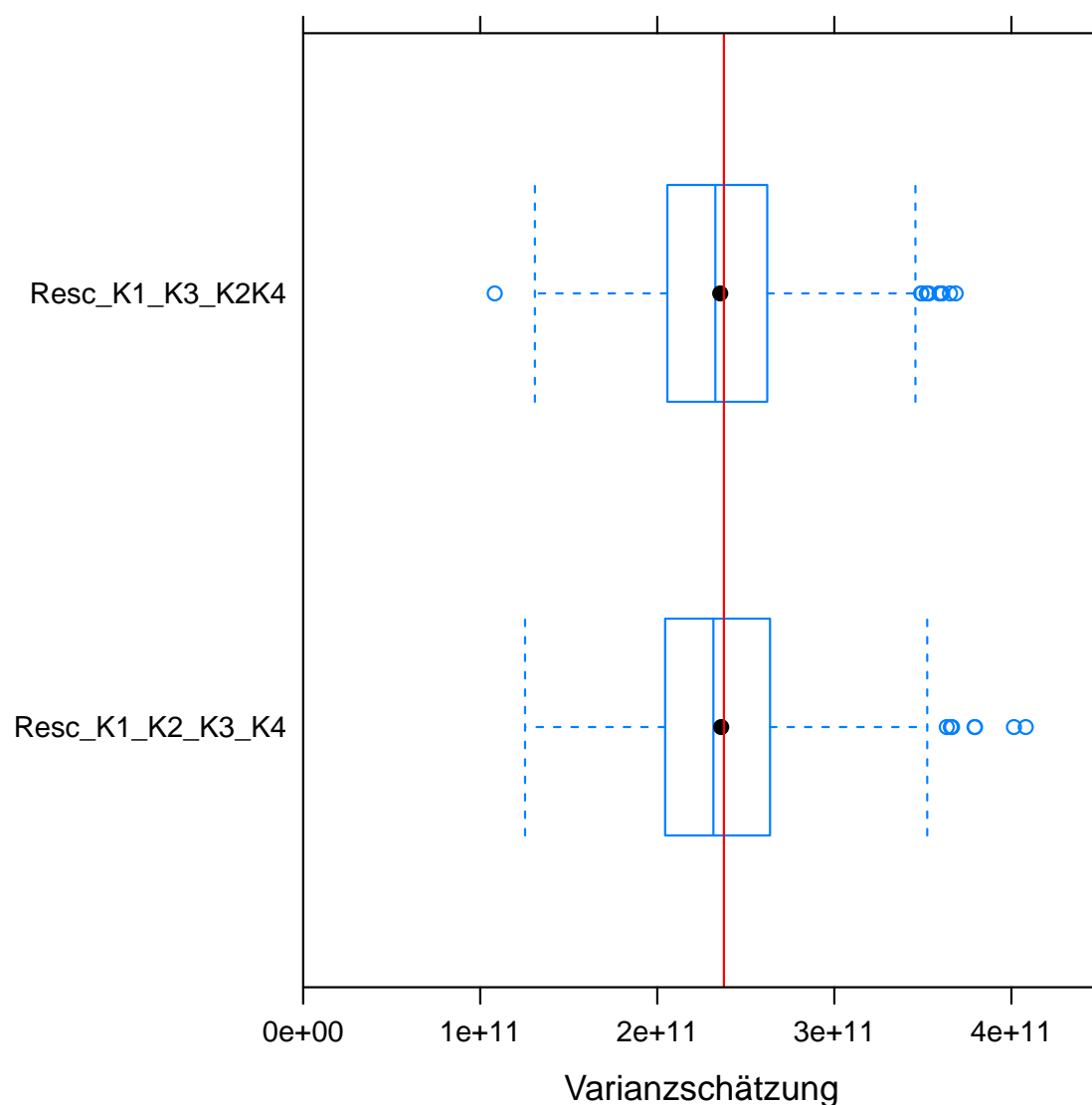


Abbildung A.3: Vergleich beim zweistufigen Stichprobendesign, Regressionsimputation, kleinen Auswahlätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

Die Abbildungen A.3 und A.4 zeigen die Ergebnisse für die Schätzer $Resc_{K1_K2_K3_K4}$ und $Resc_{K1_K3_K2K4}$ bzw. $Resc_{K1_K2_K3_K4_KF}$ und $Resc_{K1_K3_K2K4_KF}$ im Fal-

le der Anwendung von Regressionsimputation.² Es ist zu sehen, dass die Schätzer sowohl bei kleinen wie großen Auswahlsätzen auch für diese Imputationsmethode unverzerrt sind.

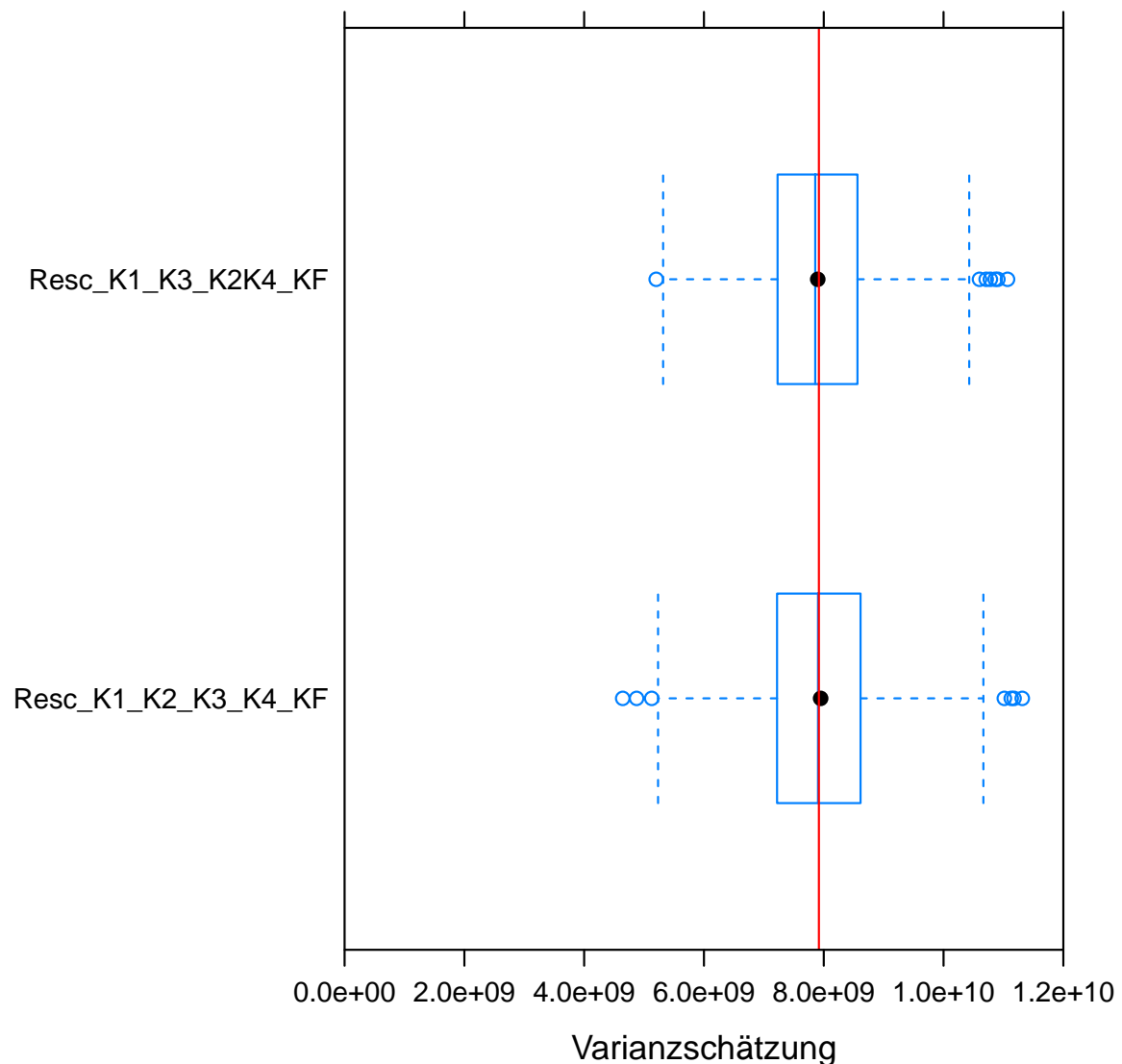


Abbildung A.4: Vergleich beim zweistufigen Stichprobendesign, Regressionsimputation, großen Auswahlsätzen, heterogenen PSU, einer gleichen PSU-Größe, MAR und Entstehung variabler fehlender Werte in der Grundgesamtheit

² Es gilt hier die Grundgesamtheit aus 5.1.2 des zweistufigen Stichprobendesigns mit einer Stratifizierung auf der ersten Stufe. Des Weiteren liegt die Parameterbeschreibung aus 5.1.5 zugrunde, wenn der Ausfall nach der ersten metrischen Hilfsvariable erzeugt wird.

Literaturverzeichnis

- Alfons, A., Filzmoser, P., B., H., J.-P., K., S., K., Münnich, R. und Templ, M. (2011):** *Synthetic Data Generation of SILC Data*. Technischer Bericht, AMELI deliverable D6.2.
URL <http://ameli.surveystatistics.net/>
- Beaumont, J.-F. und Bocci, C. (2009):** *Variance estimation when donor imputation is used to fill in missing values*. In: The Canadian Journal of Statistics, 37 (3), S. 400 – 416.
- Berger, Y. G., Bjørnstad, J., Zhang, L.-C. und Skinner, C. (2004):** *Imputation and Non-response*. Technischer Bericht, Dacseis Project, Workpackage 11, Deliverable 11.1.
URL https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Dacseis_Deliverables/DACSEIS-D11-1.pdf
- Bienas, J. L., Kott, P. und Evans, D. A. (2003):** *Applying the Delete-a-Group Jack-knife Variance Estimator to Analyses of Data from a Complex Longitudinal Survey*. 2003 Joint Statistical Meetings - Section on Survey Research Methods.
- Bihler, W., Bjørnstad, J., Davison, A., Haslinger, A., Knottnerus, P., Laaksonen, S., Münnich, R., Ohly, D., Schürle, J., Wiegert, R., Oetliker, U., Renfer, J.-P., Quatember, A., Skinner, C. und Berger, Y. G. (2004):** *Data Quality in Complex Surveys*. Technischer Bericht, Dacseis Project. Workpackage 1. Deliverable 1.1.
URL https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Dacseis_Deliverables/DACSEIS-D1-1.pdf
- Bjørnstad, J. (2007):** *Non-bayesian multiple Imputation*. In: Journal of Official Statistics, 23 (4), S. 433 – 452.
- Brick, J. M., Jones, M. E., Kalton, G. und Valliant, R. (2005):** *A simulation study of three methods of variance estimation with hot deck imputation*. In: Survey Methodology, 31, S. 151–160.
- Bruch, C. (2010):** *Varianzschätzung bei mehrstufigen Designs*. Diplomarbeit, Universität Trier.
- Bruch, C., Münnich, R. und Zins, S. (2011):** *Variance Estimation for Complex Surveys*. Technischer Bericht, AMELI deliverable D3.1.
URL <http://ameli.surveystatistics.net/>
- Chauvet, G., Deville, J.-C. und Haziza, D. (2011):** *On balanced random imputation in surveys*. In: Biometrika, 98, S. 459–471.
- Chen, J. und Shao, J. (2000):** *Nearest neighbor imputation for survey data*. In: Journal of Official Statistics, 16 (2), S. 113 – 131.

- Chen, J.** und **Shao, J.** (2001): *Jackknife variance estimation for nearest-neighbor imputation*. In: Journal of the American Statistical Association, S. 260–269.
- Chipperfield, J.** und **Preston, J.** (2007): *Efficient bootstrap for business surveys*. In: Survey Methodology, 33, S. 167–172.
- Cochran, W. G.** (1977): Sampling Techniques. New York: Wiley.
- David, M., Little, R., Samuhel, M.** und **Triest, R.** (1986): *Alternative methods for CPS income imputation*. In: Journal of the American Statistical Association, 81, S. 29–41.
- Davison, A., Hinkley, D. V.** und **Schechtman, E.** (1986): *Efficient bootstrap simulation*. In: Biometrika, 73, S. 555–566.
- Davison, A. C.** und **Sardy, S.** (2004a): *Resampling Methods for Variance Estimation*. Technischer Bericht, Dacseis Project. Workpackage 5. Deliverable 5.1.
URL https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Dacseis_Deliverables/DACSEIS-D5-1.pdf
- Davison, A. C.** und **Sardy, S.** (2004b): *Resampling Methods for Variance Estimation*. Technischer Bericht, Dacseis Project. Workpackage 5. Deliverable 5.2.
URL https://www.uni-trier.de/fileadmin/fb4/projekte/SurveyStatisticsNet/Dacseis_Deliverables/DACSEIS-D5-2.pdf
- Davison, A. C.** und **Sardy, S.** (2007): *Resampling variance estimation in surveys with missing data*. In: Journal of Official Statistics, 23, S. 371–386.
- Demnati, A.** und **Rao, J. N. K.** (2004): *Linearization variance estimators for survey data*. In: Survey Methodology, 30, S. 17–34.
- Deville, J.-C.** und **Särndal, C.-E.** (1994): *Variance estimation for the regression imputed Horvitz-Thompson estimator*. In: Journal of Official Statistics, 10, S. 381–394.
- Efron, B.** (1982): The Jackknife, the Bootstrap, and Other Resampling Plans. Philadelphia: SIAM.
- Enderle, T., Münnich, R.** und **Bruch, C.** (2013): *On the impact of response patterns on survey estimates from access panels*. In: Survey Research Methods, 7, S. 91–101.
URL <https://ojs.ub.uni-konstanz.de/srm/article/view/5036>
- Eurostat** (2013): *Handbook on Precision Requirements and Variance Estimation for ESS Household Surveys*. Technischer Bericht.
URL <http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF>
- Fay, R. E.** (1991): *A Design-Based Perspective on Missing Data Variance*. In: Proceedings of the 1991 Annual Research Conference, U.S. Bureau of the Census, S. 429–440.
- Fay, R. E.** (1996): *Alternative paradigms for the analysis of imputed survey data*. In: Journal of the American Statistical Association, 91, S. 490–498.
- Fuller, W.** und **Kim, J. K.** (2005): *Hot deck random imputation for the response model*. In: Survey Methodology, 31, S. 139–149.

- Ganninger, M. (2010):** Design Effects: Model-Based versus Design-based Approach. Gesis Series.
- Graham, R. L., V., H. D., John, P. M. und Shi, S. (1990):** *Balanced design of bootstrap simulations*. In: Royal Statistical Society Series B, 52, S. 185–202.
- Haziza, D. (2009):** *Imputation and Inference in the Presence of Missing Data*. In: **Pfaffer-fermann, D. und Rao, C.** (Hg.) Handbook of Statistics 29A: Sample Surveys: Design, Methods and Applications, Kapitel 10, S. 215–246, Amsterdam: Elsevier BV.
- Haziza, D. (2010):** *Resampling methods for variance estimation in the presence of missing survey data*. Proceedings of the Annual Conference of the Italian Statistical Society. URL <http://new.sis-statistica.org/wp-content/uploads/2013/09/>
- Horvitz, D. und Thompson, D. J. (1952):** *A generalization of sampling without replacement from a finite universe*. In: Journal of the American Statistical Association, 47, S. 663–685.
- Kalton, G. und Kish, L. (1984):** *Some efficient random imputation methods*. In: Communications in Statistics- Theory and Methods, 13, S. 1919–1934.
- Kim, J. K. (2001):** *Variance estimation after imputation*. In: Survey Methodology, 27, S. 75 – 83.
- Kim, J. K. und Fuller, W. (2004):** *Fractional hot deck imputation*. In: Biometrika, 91, S. 559–578.
- Kim, J. K. und Rao, J. N. K. (2009):** *A unified approach to linearization variance estimation from survey data after imputation for item nonresponse*. In: Biometrika, S. 917–932.
- Kim, J. K. und Shao (2014):** Statistical Methods for Handling Incomplete Data. Boca Raton: CRC Press.
- Kish, L. (1965):** Survey Sampling. John Wiley and Sons.
- Kolb, J.-P. (2012):** Methoden zur Erzeugung synthetischer Simulationsgesamtheiten. Dissertation, Universität Trier.
- Kott, P. (1998):** *Using the Delete-a-Group Jackknife Variance Estimator in Practise*. Proceedings of the Annual Meeting of the American Statistical Association, Section on Survey Research Methods.
- Kott, P. (2001):** *The delete-a-group jackknife*. In: Journal of Official Statistics, 17 (4), S. 521 – 526.
- Landerman, L. R., Land, K. C. und Pieper, C. F. (1997):** *An empirical evaluation of the predictive mean matching method for imputing missing values*. In: Sociological Methods & Research, 26 (3), S. 3 – 33.
- Lee, H., Rancourt, E. und Särndal, C.-E. (1994):** *Experiments with variance estimation from survey data with imputed values*. In: Journal of Off, 10, S. 231 – 243.

- Little, R. (1986):** *Missing Data in Census Bureau Surveys*. In: Proceedings of the Second Annual Census Bureau Research Conference, Washington, DC: U.S. Department of Commerce, Bureau of the Census.
- Little, R. und Rubin, D. (2002):** Statistical analysis with missing data. Wiley series in probability and mathematical statistics. Probability and mathematical statistics, Wiley.
- Lohr, S. (1999):** Sampling: Design and Analysis. Pacific Grove: Duxbury Press.
- Longford (2005):** Missing Data and Small-Area Estimation. Springer.
- Mashreghi, Z., Leger, C. und Haziza, D. (2014):** *Bootstrap methods for imputed data from regression, ratio and hot-deck imputation*. In: The Canadian Journal of Statistics, 42 (1), S. 142 – 167.
- Matei, A. und Tillé (2005):** *Evaluation of variance approximations and estimators in maximum entropy sampling with unequal probability and fixed sample size*. In: Journal of Official Statistics, 21, S. 543–570.
- McCarthy, P. J. (1966):** *Replication: An Approach to the Analysis Data from Complex Surveys*. Technischer Bericht, Public Health Service Publication 14, National Center for Health Statistics, Public Health Service, Washington D.C: U.S. Government Printing Office, Series 2.
- McCarthy, P. J. (1969):** *Pseudo-replication: half samples*. In: Review of the International Statistical Institute, 37, S. 239–264.
- Münnich, R. (2008):** *Varianzschätzung in komplexen Erhebungen*. In: Austrian Journal of Statistics, 37, S. 319–334.
- Münnich, R., Gabler, S., Bruch, C., Burgard, J. P., Enderle, T., J.-P., K. und T., Z. (2015):** *Tabellenauswertungen im Zensus unter Berücksichtigung fehlender Werte*. In: AStA Wirtschafts- und Sozialstatistisches Archiv, 9 (3), S. 269 – 304.
- Nigam, A. K. und Rao, J. (1996):** *On balanced bootstrap for stratified multistage samples*. In: Statistica Sinica, 6, S. 199–214.
- Preston, J. (2009):** *Rescaled bootstrap for stratified multistage sampling*. In: Survey Methodology, 35, S. 227–234.
- R Core Team (2015):** R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL <https://www.R-project.org/>
- Rao, J. (1990):** *Variance Estimation Under Imputation for Missing Data*. Technischer Bericht, Statistics Canada.
- Rao, J. und Shao, J. (1992):** *Jackknife variance estimation with survey data under hot deck imputation*. In: Biometrika, 79 (4), S. 811– 822.
- Rao, J. und Wu, C. F. J. (1988):** *Resampling inference with complex survey data*. In: Biometrika, 83 (401), S. 231 – 241.

- Rao, J. N. K. und Shao, J. (1996):** *On balanced half-sample variance estimation in stratified random sampling.* In: Journal of the American Statistical Association, 91, S. 343–348.
- Rao, J. N. K., Wu, C. F. J. und Yue, K. (1992):** *Some recent work on resampling methods for complex surveys.* In: Survey Methodology.
- Rubin, D. (1986):** *Basic ideas of multiple imputation for nonresponse.* In: Survey Methodology, 12, S. 37 – 47.
- Rubin, D. B. (1987):** Multiple Imputation for Nonresponse in Surveys. Wiley.
- Saigo, H., Shao, J. und Sitter, R. (2001):** *A repeated half-sample bootstrap and balanced repeated replications for randomly imputed data.* In: Survey Methodology, 27 (2), S. 189–196.
- Särndal, C.-E. (1992):** *Methods for estimating the precision of survey estimates when imputation has been used.* In: Survey Methodology, 18 (2), S. 241– 252.
- Särndal, C.-E. und Lundström, S. (2005):** Estimation in Surveys with Nonresponse. New York: Wiley.
- Särndal, C.-E., Swensson, B. und Wretman, J. (1992):** Model Assisted Survey Sampling. New York: Springer.
- Schwender, H. und Fritsch, A. (2013):** *scrim: Analysis of High-Dimensional Categorical Data such as SNP Data.* R package version 1.3.3.
URL <https://CRAN.R-project.org/package=scrim>
- Schwender, H. und Ickstadt, K. (2008):** *Imputing missing genotypes with weighted k nearest neighbors.* Technischer Bericht, University of Dortmund.
- Shao, J. (2002):** *Replication Methods for Variance Estimation in Complex Surveys with Imputed Data.* In: Groves, R. M., A., D. D., Eltinge, J. L. und Little, R. J. A. (Hg.) Survey Nonresponse, New-York: Wiley.
- Shao, J., Chen, Y. und Chen, Y. (1998):** *Balanced repeated replication for stratified multistage survey data under imputation.* In: Journal of the American Statistical Association, 93 (442), S. 819– 831.
- Shao, J. und Sitter, R. (1996):** *Bootstrap for imputed survey data.* In: Journal of the American Statistical Association, 91 (435), S. 1278– 1288.
- Shao, J. und Steel, P. (1999):** *Variance estimation for survey data with composite imputation and nonnegligible sampling fractions.* In: Journal of the American St, 94 (445), S. 254 – 265.
- Shao, J. und Tang, Q. (2011):** *Random group variance estimators for survey data with random hot deck imputation.* In: Journal of Official Statistics, 27, S. 507–526.
- Shao, J. und Tu, D. (1995):** The jackknife and bootstrap. New-York: Springer.
- Shao, J. und Wu, C. F. J. (1989):** *A general theory for jackknife variance estimation.* In: The Annals of Statistics, 17, S. 1176–1197.

- Templ, M., Alfons, A., Kowarik, A. und Prantner, B. (2016):** VIM: Visualization and Imputation of Missing Values. R package version 4.5.0.
URL <https://CRAN.R-project.org/package=VIM>
- Valliant, R. (1987):** *Some prediction properties of balanced half-sample variance estimators in single-stage sampling.* In: Journal of the Royal Statistical Society Series B, 49, S. 68–81.
- Wolter, K. M. (2007):** Introduction to Variance Estimation. New York: Springer.