

Die Gewichtung des Mikrozensus-Panels
Ein Vergleich von Kalibration und Propensity Score Adjustment

Berlin 2018

Inaugural-Dissertation zur Erlangung des akademischen Grades eines Doktors
der Wirtschaftswissenschaft des Fachbereichs Wirtschaftswissenschaft der
Freien Universität Berlin

vorgelegt von: Ivo Marek

Erstgutachter: Prof. Dr. Ulrich Rendtel
(Freie Universität Berlin)

Zweitgutachter: Prof. Dr. Martin Spieß
(Universität Hamburg)

Tag der Disputation: 11. Februar 2019

Inhaltsverzeichnis

1	Einleitung	4
2	Grundlagen	7
2.1	Grundbegriffe der Stichprobentheorie	7
2.2	Quasi-Randomization	9
2.3	Asymptotische Betrachtungsweise	12
2.4	Gleichmäßige Konvergenz	21
3	Kalibration	29
3.1	Kalibration bei Nonresponse	29
3.2	Varianz des Kalibrationsschätzers	32
3.3	Im Vergleich zu Lundström und Särndal (1999)	35
4	Propensity Score Adjustment	37
4.1	Score-Funktion und Fisher-Information	38
4.2	Existenz der Maximum-Likelihood-Lösung	40
4.3	Approximation der Maximum-Likelihood-Lösung	43
4.4	Varianz des PSA-Schätzers	45
4.5	Kalibration des PSA-Schätzers	47
4.6	Die Varianz bei Abhängigkeiten	49
5	Simulation	52
5.1	Das Erhebungsdesign des Mikrozensus	52
5.2	Das Sozio-oekonomische Panel	53
5.3	Theoretische Herleitung der Simulation	53
5.4	Ergebnisse der Simulation	60
6	Zusammenfassung	68
A	Anhang	69
A.1	Eine untere Schranke für $\hat{\mathbf{T}}_{\text{NR}}$	69
A.2	Nicht-natürliche Link-Funktionen	69
A.3	Approximation der GEE Lösung	71
A.4	Bootstrap des PSA-Schätzers	73
B	Englische Zusammenfassung	75
	Literatur	77

1 Einleitung

In den Wirtschafts- und Sozialwissenschaften kommt der Darstellung und Erklärung des sozialen Wandels eine zentrale Bedeutung zu. Im Vordergrund stehen dabei insbesondere die Veränderungen der ökonomischen Situation der privaten Haushalte, Inklusions- und Exklusionsprozesse am Arbeitsmarkt sowie der Wandel der sozialen Beziehungen. Klassische Beispiele sind: Welche Folgen haben veränderte Einstellungen zu Ehe und Familie auf die zukünftige demographische Entwicklung der Gesellschaft? Von welchen Faktoren hängt die Erwerbsbeteiligung von Frauen ab, und wie wird sie sich zukünftig verändern? Welche Erklärungskraft haben traditionelle sozialstrukturelle Unterscheidungen etwa für die Bildungsbeteiligung? Diese Liste von Fragen ließe sich beliebig verlängern. Ihnen allen ist gemeinsam, dass ihre Beantwortung empirische Informationen voraussetzt, die in vergleichbarer Form über einen längeren Zeitraum erhoben wurden.

Dabei spielen empirische Daten, die über ein Panel-Design, die sogenannte Panel-Erhebungen, gewonnen werden eine wichtige Rolle. Der Begriff Panel-Erhebungen bezeichnet dabei eine periodisch wiederholte Befragung derselben Untersuchungseinheiten. Sie dienen dazu Veränderungen im Zeitverlauf zu ermitteln, d.h. es ist möglich eindeutig festzustellen, welche Individuen (Haushalte) tatsächlich vom sozialen Wandel betroffen sind, der sich im Verlauf der Zeit beobachten lässt. Allerdings sind bei dieser Art der Befragung die felddingten Ausfälle von Teilnehmern des Panels problematisch. Diese Ausfälle treten auf, weil Personen (Haushalte) in der folgenden Welle nicht mehr erreicht werden oder weil sie die weitere Teilnahme am Panel verweigern. Als Konsequenz dieser Panelausfälle beobachtet man in vielen Panels das Phänomen, dass nach einigen Jahren der Befragung ein großer Teil der ursprünglichen Stichprobenmitglieder ausgefallen ist. So haben von 1968 bis 1975 ungefähr ein Viertel der ursprünglichen Stichprobenmitglieder die „Panel Study of Income Dynamics (PSID)“ verlassen, Fitzgerald et al. (1998). Im Sozio-ökonomischen Panel (SOEP) schieden nach den ersten acht Jahren der Befragung ca. 34% der ursprünglichen Stichprobenmitglieder (aller Verläufe erhebungsbedingt) aus, Kroh und Spiess (2006).

Durch Panelausfälle verringert sich zunächst die Fallzahl der Stichprobe, sodass sich die Effizienz der Schätzer der Grundgesamtheitsparameter im Vergleich zu einer Stichprobe ohne Panelausfälle verringert. Zudem besteht die Gefahr der Verzerrung der Analyseergebnisse, wenn sich Teilnehmer von Nichtteilnehmern systematisch unterscheiden.

Der Mikrozensus (MZ) ist EU-weit die größte Haushaltsstichprobe und wird seit 1957 durchgeführt. Mit dem Mikrozensusgesetz 1985 (Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt vom 10. Juni 1985) besteht eine rechtliche Grundlage, dass Befragungen in vier aufeinanderfolgenden Jahren durchgeführt werden dürfen. Allerdings ist es erst seit dem Mikrozensusgesetz 1996 (Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt vom 17. Januar 1996) möglich, die Querschnittserhebungen des Mikrozensus zu einem Paneldatensatz zusammenzuführen (MZ-Panel). Da die Personen bzw. Haushalte vier Jahre lang hintereinander befragt werden, ist es möglich Kurzzeitpanels zu konstruieren. Die Attraktivität in der Nutzung des Mikrozensus für Längsschnittanalysen liegt zum einen in der hohen Fallzahl (200'000 Auskunftspersonen) und zum anderen in der Teilnahmepflicht der ausgewählten Personen bzw. Haushalte.

Allerdings stehen diesen beiden großen Vorteilen des Mikrozensus auch einige Nachteile gegenüber. Das Hauptproblem bei der Längsschnitts-Auswertung des Mikrozensus besteht in der Tatsache, dass

der Mikrozensus als Flächenstichprobe erhoben wird. So werden Haushalte und Personen nicht direkt aus der Grundgesamtheit für die Stichprobe gezogen, sondern die Auswahlinheiten, sogenannte Auswahlbezirke, werden über räumliche Zuordnungen gebildet. Dies hat zur Folge, dass Haushalte, die aus einem zu befragenden Auswahlbezirk ziehen, nicht weiterverfolgt werden. Sie werden bei der nächsten Erhebung nicht mehr befragt und ihre Angaben stehen somit für Längsschnittanalysen nicht weiter zur Verfügung. Dagegen werden Haushalte und Personen, die in einen zu befragenden Auswahlbezirk ziehen, bei der nächsten Befragung miterhoben. Auch von diesen Personen existieren keine vollständigen Angaben über den gesamten Mikrozensus-Längsschnitt. Dieser Informationsverlust, der durch Zuzug oder Fortzug einer Person oder eines Haushalts entsteht, kann wegen der Betrachtung von nur räumlich immobilen Personen zu einem Bias führen. Es sei beispielsweise das interessierende Merkmal der Übergang aus der Arbeitslosigkeit in die Erwerbstätigkeit. Wenn Personen, die diesen Übergang vollziehen in verstärktem Maße mobil sind, so wird man diesen Übergang auf Basis der räumlich immobilen Personen unterschätzen. In der Tat gibt es empirische Hinweise für den Mikrozensus, Bašić (2008), dass ein Wechsel aus der Arbeitslosigkeit in die Erwerbstätigkeit ein erhöhtes Mobilitätsrisiko impliziert.

Die Auswertung eines Panels kann im Design- bzw. Modell-basierten Ansatz erfolgen. Der Design-basierte Ansatz ist die Standardauswertungsmethode in der amtlichen Statistik. Bei diesem Ansatz ist man an der Schätzung von Gesamtwerten (Totals) und Anteilswerten in endlichen Grundgesamtheiten interessiert. Es geht beispielsweise um die Frage, wie viele Personen im Erhebungsgebiet in einem bestimmten Zeitintervall kontinuierlich arbeitslos gewesen sind oder es soll ermittelt werden, wie hoch der Anteil der Personen in der Bevölkerung ist, die bei dem Start des Panels arbeitslos waren und nach einer bestimmten Zeit erwerbstätig geworden sind. Dieser Ansatz unterstellt, dass alle Merkmale der Populationsmitglieder unbekannte, feste Parameter der Grundgesamtheit sind. Hierbei wird nur die Ziehung der Stichprobenmitglieder aus der Grundgesamtheit als zufällig betrachtet. Die zu schätzenden Gesamtwerte sind damit Summen der unbekannt Parameterwerte. Im Gegensatz dazu werden beim Modell-basierten Ansatz Merkmale von Populationsmitgliedern als zufällige Größen betrachtet. Man interessiert sich dabei an einer modellhaften Beschreibung von Zusammenhängen zwischen verschiedenen Merkmalen. Es geht beispielsweise um die Frage, wie wahrscheinlich ist es, dass eine arbeitslose Frau mit bestimmten Eigenschaften wieder erwerbstätig wird. Bei diesem Ansatz ist das Modell durch eine feste, in der Regel kleine Zahl von unbekannt Parametern bestimmt, die auf Basis der Stichprobe geschätzt werden.

Die vorliegende Arbeit verfolgt bei der Auswertung des MZ-Panels den Design-basierten Ansatz und beschäftigt sich mit der Fragestellung, ob mit einem Gewichtungsansatz Auswertungen auf Basis des MZ-Panels trotz fehlender Information über räumlich mobilen Personen valide Aussagen über die Population gestatten, aus der die Stichprobe gezogen wurde, d.h. über die Gesamtbevölkerung Deutschlands.

Die gebundene Hochrechnung oder auch im Folgenden Kalibration genannt ist ein Standardverfahren der amtlichen Statistik, das auch in Abwesenheit von Antwortausfällen (Nonresponse) den Standardfehler reduzieren kann. Dabei erfolgt die Schätzung der interessierenden Gesamtwerte unter Zuhilfenahme der Verteilung von bekannten Hilfsmerkmalen in der Population. Ein Beispiel für ein Hilfsmerkmal ist das Alter. Aus der Bevölkerungsfortschreibung ist die Anzahl der Personen innerhalb einer jeden Altersgruppe in der Bevölkerung bekannt.

Kalibration kann auch im Falle von Nonresponse angewendet werden. Die Idee dabei ist, dass

eine durch Nonresponse hervorgerufene Verzerrung damit entgegengewirkt wird. Für den Fall der räumlichen Mobilität besteht also die Hoffnung, dass systematische Verzerrungen durch Kalibration reduziert werden. Der Kalibrationsansatz bei Nonreponse wird in Kapitel 3 dargestellt. Die Herleitung der Varianz basiert im Gegensatz zu Lundström und Särndal (1999) auf dem asymptotischen Konzept von Fuller und Isaki (1981) und Isaki und Fuller (1982).

Im Vordergrund des asymptotischen Konzepts steht die Design-Konsistenz. Ein Schätzer ist Design-konsistent für den Populationsmittelwert, falls deren Differenz in Wahrscheinlichkeit bezüglich des Ziehungsverfahrens gegen Null konvergiert. Hierbei wächst die Stichprobe zusammen mit der Grundgesamtheit. Dies erlaubt für nicht lineare Schätzer eine Aussage darüber wie weit der Schätzer im Sinne eines Wahrscheinlichkeitsmaßes vom Populationswert entfernt ist. Erweitert man das Konzept um die gleichgradige Integrierbarkeit, so lässt sich die Design-Konsistenz auf die Konvergenz im p -ten Mittel übertragen. Dies wird im Kapitel 2 dargestellt.

Eine Alternative zur Kalibration ist, die Wahrscheinlichkeiten für räumliche Immobilität mit Hilfe eines statistischen Modells zu schätzen und jede räumlich immobile Person mit dem reziproken Wert ihrer geschätzten Wahrscheinlichkeit immobil zu bleiben zu gewichten. Die Idee dabei ist, den Personen mit hoher Mobilitätsneigung, die aber räumlich immobil geblieben sind, ein hohes Gewicht zu geben um für all diejenigen zu kompensieren, die tatsächlich mobil waren. Dieser Ansatz wird Propensity Score Adjustment (PSA) genannt und der darauf basierende Schätzer PSA-Schätzer genannt. Der Ansatz wird in Kapitel 4 dargestellt.

Die Bewertung des Kalibrationsansatzes und des Propensity Score Adjustments für die Reduktion der Verzerrung aufgrund von räumlicher Mobilität auf die Schätzung der Gesamtwerte kann nicht auf der Grundlage des MZ überprüft werden. Ebenso kann das Ausmaß der Ausfälle durch räumliche Mobilität auf die Schätzung der Standardfehler beurteilt werden. Aus diesem Grund wurden die Daten des SOEPs herangezogen, das räumliche Mobilität erfasst. Die Performance der beiden Ansätze wird am Beispiel der Modellierung der Arbeitsmarktdynamik untersucht. Hierbei werden die Übergänge zwischen den verschiedenen Erwerbszuständen betrachtet. Die Beurteilung der beiden Verfahren erfolgt bzgl. des relativen Biases bei der Schätzung der Erwerbsübergänge. Zusätzlich werden die beiden Verfahren anhand der geschätzten Standardfehler beurteilt. Dabei wird eine nicht-parametrische Simulationsstudie mit dem SOEP durchgeführt. Hierzu werden aus dem SOEP Stichproben gezogen. Für jede gezogene Stichprobe werden Bootstrap-Stichproben mit Zurücklegen wiederholt gezogen. Das Verfahren hat unter Berücksichtigung eines Korrekturterms eine zum Schätzer asymptotisch äquivalente Varianz. Der Korrekturterm ist indirekt von dem zu betrachteten Schätzer abhängig, aber unabhängig von den Ausgleichsgewichten des jeweiligen Schätzers. Der Vorteil dieser Vorgehensweise ist, dass Abhängigkeiten zwischen Merkmalen und der räumlichen Mobilität nicht zerstört werden und somit kein Schätzer bevorzugt wird. Falls beispielsweise ein logistisches Regressionsmodell auf Basis des SOEP geschätzt wird und mit dem logistischen Regressionsmodell die Mobilität künstlich erzeugt wird, könnte der PSA-Schätzer, der dieselben Merkmale im Modell enthält, bevorzugt sein. Die Korrektheit des Verfahrens wird in Abschnitt 5.3 für alle betrachteten Schätzer und Varianzschätzer gezeigt.

Die unterschiedlichen Ansätze und Schätzer werden mit Hilfe der nicht-parametrischen Simulationsstudie am Beispiel der Arbeitsmarktdynamik in Abschnitt 5.4 miteinander verglichen.

2 Grundlagen

Der Mikrozensus ist eine Befragung der in Deutschland lebenden Personen, die von den Statistischen Ämtern des Bundes und der Länder durchgeführt wird. Im Gegensatz zu einer Volkszählung werden nicht alle Personen, sondern nur eine zufällig ausgewählte Stichprobe befragt. Die Befragten sind zwar Personen, jedoch werden Wohnung bzw. Zusammenfassungen von Wohnungen, sogenannte Auswahlbezirke, in denen die Personen leben, zufällig ausgewählt. Alle Personen in einem zufällig ausgewählten Auswahlbezirk werden befragt.

Der Mikrozensus bietet die Möglichkeit den Anteil der Erwerbslosen zu schätzen. Dieser Anteil kann mit Hilfe des Horvitz-Thompson Schätzer, Horvitz und Thompson (1952), erwartungstreu geschätzt werden. Dieser und dessen Varianz wird in Abschnitt 2.1 dargestellt. Theoretisch lässt sich mit diesem Schätzer auch der Anteil der Erwerbslosen, die innerhalb drei Jahre wieder erwerbstätig werden, schätzen. Da aber die Information von räumlich mobilen Personen nach bzw. vor Umzug fehlt, muss der Schätzer auf den räumlich immobilen Personen und nicht auf der gesamten Stichprobe basieren, wie das beim Horvitz-Thompson Schätzer der Fall ist. Im Abschnitt 2.2 wird deshalb die Stichprobenziehung um eine zusätzliche Auswahlphase für Antwortausfälle (Nonresponse) erweitert. Der dort dargestellte Schätzer ist rein theoretischer Natur, da er für die Antwortausfälle bekannte Wahrscheinlichkeiten voraussetzt. Da die Wahrscheinlichkeiten im Allgemeinen unbekannt sind, werden Verfahren betrachtet, bei denen die unbekanntes Wahrscheinlichkeiten durch Schätzer ersetzt werden. Das führt aber zu nicht linearen Schätzern. Nicht lineare Schätzer lassen sich durch den Satz von Taylor durch einen linearen Term und ein nicht lineares Restglied ausdrücken. In den Abschnitten 2.3 und 2.4 wird der Frage nachgegangen, wann das Restglied vernachlässigt werden kann, damit der Fehler bei der Berechnung von Erwartungswerten und der Varianzen auf Basis des linearen Terms nicht allzu groß ist.

2.1 Grundbegriffe der Stichprobentheorie

Im Design-basierten Ansatz ist der Ausgangspunkt eine feste Grundgesamtheit von sogenannten Einheiten $U = \{1, 2, \dots, N\}$ der Größe N . Ein Beispiel für eine Grundgesamtheit ist die Bevölkerung Deutschlands. Hierbei kann man zwischen Haushalten und Personen als Einheiten unterscheiden. Im Zusammenhang mit den Einheiten stehen die Ausprägungen von Merkmalen. Eine beliebige Person hat zum Beispiel das Merkmal Erwerbszustand, d.h. die Person ist entweder erwerbstätig, arbeitslos oder entzieht sich dem Arbeitsmarkt. Ein bestimmtes Merkmal ist die Untersuchungsvariable und wird mit y bezeichnet. Sie nimmt für die Einheit $k \in U$ die Ausprägung y_k an. Ein mögliches Auswertungsproblem besteht darin den Mittelwert oder den Gesamtwert aller Werte der Untersuchungsvariable einer Grundgesamtheit anhand einer Stichprobe zu schätzen. Angenommen man interessiert sich für den Anteil der arbeitslosen Personen in Deutschland. Aus Kosten- und Zeitgründen können aber nicht alle in Deutschland lebenden Personen nach ihrem Erwerbszustand befragt werden. Welche Personen sollen ausgewählt werden, und wie schätzt man mittels der erhobenen Daten den Anteil der arbeitslosen Personen? Eine Menge $s \subseteq U$ heißt Stichprobe.¹ Beispielsweise kann man auf Basis des Mikrozensus den Anteil der arbeitslosen Personen in Deutschland schätzen. Entscheidend für die Schätzung der Mittelwerte ist, auf welche Art und Weise die Einheiten aus der Grundgesamtheit in die Stichprobe gelangen. Dieser Prozeß der Auswahl der Einheiten aus der Grundgesamtheit wird Stichprobende-

¹ U selbst und die leere Menge $\{\}$ sind Stichproben.

sign genannt. Das Stichprobendesign p ist definiert auf der Menge aller möglichen Stichproben, die aus der Grundgesamtheit gezogen werden können. Nach diesem Design wird jeder Stichprobe s aus der Menge der möglichen Stichproben $\mathcal{S} = \{s \subseteq U\}$ eine Wahrscheinlichkeit $p(s) \geq 0$ zugeordnet, so dass $\sum_{\mathcal{S}} p(s) = 1$ gilt.² Die Kenntnis des Stichprobendesigns spielt eine entscheidende Rolle um Größen der Grundgesamtheit erwartungstreu schätzen zu können. Formal ist das Design eine Funktion $p: \Omega \rightarrow [0, 1]$ wobei $\Omega = \mathcal{S}$ ist.

Auf Basis des Stichprobendesigns lässt sich die Auswahlwahrscheinlichkeit für jede Einheit $k \in U$ bestimmen. Hierfür wird der Begriff des Ereignisses benötigt. Eine Menge $A \subseteq \Omega$ heißt Ereignis. $\mathcal{H} = \{A \subseteq \Omega\}$ ist die Menge aller Ereignisse. Für Ereignisse $A \in \mathcal{H}$ lassen sich Wahrscheinlichkeiten mit Hilfe des Designs berechnen: $P(A) = \sum_A p(s)$. Für $A = \{\}$ ist $\sum_A p(s) = 0$ zu verstehen.³ Zur Bestimmung von Auswahlwahrscheinlichkeiten für Einheiten $k \in U$ sind die folgenden Ereignisse nützlich. Im Folgenden bezeichnet $\{k \in s\}$ das Ereignis, dass die Einheit $k \in U$ in der Stichprobe enthalten ist. Die Wahrscheinlichkeit für dieses Ereignis ist $P(\{k \in s\}) = \sum_{\{k \in s\}} p(s)$ und wird mit π_k bezeichnet. Es wird davon ausgegangen, dass $\pi_k > 0$ für alle $k \in U$ ist. Somit hat jede Einheit eine positive Wahrscheinlichkeit ausgewählt zu werden. Mit $\{k, l \in s\}$ wird das Ereignis bezeichnet, dass die Einheiten $k \in U$ und $l \in U$ in der Stichprobe vertreten sind. Die Wahrscheinlichkeit dieses Ereignisses ist $\pi_{kl} = P(\{k, l \in s\}) = \sum_{\{k, l \in s\}} p(s)$, wobei $\pi_{kk} = \pi_k$ ist. Wiederum sei $\pi_{kl} > 0$ für alle $k, l \in U$ vorausgesetzt.

Weiterhin wird eine Zufallsgröße benötigt, die angibt, ob ein Ereignis eintritt oder nicht. $1_A: \Omega \rightarrow \mathbb{R}$ sei die Indikatorfunktion für ein Ereignis $A \in \mathcal{H}$ mit $1_A(s) = 1$, falls $s \in A$ ist und $1_A(s) = 0$, falls das Gegenteil der Fall ist. Der Erwartungswert von 1_A ist $E_p[1_A] = P(A)$. Hierbei wird der Erwartungswert E_p , die Varianz V_p , bzw. Kovarianz Cov_p unter dem Design p bestimmt.⁴ Für die Ereignisse $\{k \in s\}$ und $\{k, l \in s\}$ lassen sich die Erwartungswerte berechnen: $\pi_k = E_p[1_{\{k \in s\}}]$ und $\pi_{kl} = E_p[1_{\{k, l \in s\}}]$. Damit lässt sich auch die erwartete Stichprobengröße berechnen: $n = \sum_U E_p[1_{\{k \in s\}}] = \sum_U \pi_k$.

Mit der obigen Notation lässt sich der Horvitz-Thompson Schätzer, Horvitz und Thompson (1952), $\hat{Y}_{\text{HT}}: \Omega \rightarrow \mathbb{R}$ angeben. Der Schätzer basiert auf der gewichteten Stichprobe. Das Konzept, dass eine Einheit für viele Einheiten in der Grundgesamtheit repräsentativ ist, ist eine natürliche Konsequenz aus dem Ziehungsverfahren, um Informationen bezüglich der Grundgesamtheit zu erhalten. Im Fall der Schätzung eines Mittelwertes $Y = N^{-1} \sum_U y_k$ ist der Horvitz-Thompson Schätzer der Mittelwert der mit $w_k = 1/\pi_k$ gewichteten y_k :

$$\hat{Y}_{\text{HT}} = \frac{1}{N} \sum_U 1_{\{k \in s\}} w_k y_k \quad (2.1)$$

Die Gewichte w_k , $k \in U$, heißen Design-Gewichte. Dieser Schätzer hat eine wünschenswerte Eigenschaft, die seine Beliebtheit rechtfertigt, nämlich er ist erwartungstreu unter dem Stichprobendesign p :

$$E_p[\hat{Y}_{\text{HT}}] = \frac{1}{N} \sum_U E_p[1_{\{k \in s\}}] w_k y_k = Y$$

Allerdings sichert die Tatsache, dass \hat{Y}_{HT} erwartungstreu ist, nicht vor der Möglichkeit, im Einzelfall

²Für eine Menge $A \subseteq U$ bedeutet \sum_A , dass über alle Elemente der Menge A summiert wird, $\sum_{k \in A}$.

³Wird ein Ereignis über eine Aussage A beschrieben, d.h. $\{s \in \mathcal{S} : A(s)\}$ wobei die Aussage für die Stichprobe entweder richtig oder falsch ist, dann wird im Folgenden nur die Aussage ausgeschrieben, d.h. $\{A(s)\}$.

⁴Der Index p wird hier verwendet um die Berechnung von der Berechnung in 2.3 abzugrenzen.

Schätzergebnisse zu erhalten, die sehr weit von dem wahren Mittelwert Y liegen können.⁵ Deswegen ist es von besonderer Bedeutung auch die Varianzen für die Schätzergebnisse angeben zu können. Für den Horvitz-Thompson Schätzer lässt sich die Varianz wie folgt berechnen:

$$V_p[\hat{Y}_{\text{HT}}] = \frac{1}{N^2} \sum \sum_U \text{Cov}_p[1_{\{k \in s\}}, 1_{\{l \in s\}}] \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} = \frac{1}{N^2} \sum \sum_U \Delta_{kl} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} \quad (2.2)$$

mit $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$. Da im Allgemeinen aber nur eine Stichprobe s zur Verfügung steht, ist die obige Formel zur Berechnung der Varianz nicht anwendbar. Allerdings lässt sich die Varianz erwartungstreu durch

$$\hat{V}_{\text{HT}} = \frac{1}{N^2} \sum \sum_U 1_{\{k, l \in s\}} \frac{\Delta_{kl}}{\pi_{kl}} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} \quad (2.3)$$

schätzen, da der Erwartungswert $E_p[1_{\{k, l \in s\}}] = \pi_{kl}$ ist.

Es sei hier noch bemerkt, dass für eine fixe Stichprobengröße $\sum_U 1_{\{k \in s\}} = \sum_U \pi_k = n$, fast sicher⁶, die Varianz mit der Darstellung

$$V_p[\hat{Y}_{\text{HT}}] = -\frac{1}{2N^2} \sum \sum_U \Delta_{kl} \left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l} \right)^2 \quad (2.4)$$

berechnet werden kann. Auf dieser Darstellung basiert der erwartungstreu Varianzschätzer von Yates und Grundy (1953) und Sen (1953):

$$\hat{V}_{\text{YG}} = -\frac{1}{2N^2} \sum \sum_U 1_{\{k, l \in s\}} \frac{\Delta_{kl}}{\pi_{kl}} \left(\frac{y_k}{\pi_k} - \frac{y_l}{\pi_l} \right)^2 \quad (2.5)$$

Der Vorteil der Darstellung in (2.4) und des Varianzschätzers (2.5) ist, dass sich diese leicht nach oben abschätzen lassen, indem nur Terme mit $\Delta_{kl} < 0$ betrachtet werden.

In einer idealen Welt lassen sich der Horvitz-Thompson Schätzer und dessen Varianz schätzen. Jedoch wird die Ausprägung des Untersuchungsmerkmals y in einer realen Welt nicht für jede Einheit der Stichprobe erhoben. Das Phänomen, dass die Merkmalsausprägungen einzelner Einheiten der Stichprobe unbekannt sind, nennt man Nonresponse. Im Folgenden sei Nonresponse definiert als der Zustand, dass die Ausprägung des Untersuchungsmerkmals einer Einheit der Stichprobe unbekannt ist. Konzeptionell wird davon ausgegangen, dass ein stochastischen Prozeß für die Beobachtbarkeit des Untersuchungsmerkmal verantwortlich ist.

2.2 Quasi-Randomization

Sei ein Stichprobendesign $p_D: \mathcal{S} \rightarrow [0, 1]$ gegeben und eine Stichprobe wird mit Hilfe dieses Designs ausgewählt. Das Untersuchungsmerkmal y wird nicht für jede Einheit der Stichprobe s beobachtet. Sei $r \subseteq s$ die Menge der Einheiten, für die das Untersuchungsmerkmal beobachtet wird. Quasi-Randomization bedeutet, dass r wie eine zufällige Auswahl behandelt wird. Damit ist die theoretische Chance gegeben, einen erwartungstreuen Schätzer zu erhalten.

Bei Quasi-Randomization, Oh und Scheuren (1983), geht man also davon aus, dass die Einhei-

⁵Dies gilt insbesondere bei Stichprobendesigns mit ungleichen Auswahlwahrscheinlichkeiten und Merkmalen mit großer Streuung, vgl. Basu's berühmtes Elefanten-Beispiel, Basu (1971).

⁶Ein Ereignis $A \in \mathcal{H}$ heißt fast sicher, falls $P(A) = 1$ gilt. Im vorliegenden Fall bedeutet dies, dass alle $s \in \Omega$ mit $p(s) > 0$ eine Stichprobengröße von n haben.

ten, für die die Ausprägung von y beobachtet wird, durch einen Response-Prozeß $q(\cdot|s) : \mathcal{S} \rightarrow [0, 1]$ ausgewählt werden. Man nimmt also an, dass die Auswahl der Einheiten und deren Untersuchungsmerkmal in zwei Phasen erfolgt. Die Auswahl von Einheiten in zwei Schritten geht ursprünglich auf Neyman (1938) zurück. In der ersten Phase werden Merkmale erhoben, die mit dem Untersuchungsmerkmal korreliert sind, jedoch nicht das Untersuchungsmerkmal selbst. Erst in einer zweiten Phase wird das Untersuchungsmerkmal für eine Teilmenge der Einheiten der ersten Phase erhoben. Dieses Prinzip lässt sich auf Nonresponse übertragen. Der Response-Prozeß ist aber im Gegensatz zum Stichprobendesign unbekannt. Eine erste Restriktion für den Response-Prozeß ist $\sum_{\mathcal{S}} q(r|s) = 1$ für $s \in \mathcal{S}$ mit $p_D(s) > 0$. Für Mengen, die Einheiten enthalten, die nicht selbst durch das Stichprobendesign ausgewählt wurden, d.h. $r \not\subseteq s$, gilt $q(r|s) = 0$. Für nicht-realisierte Stichproben, $p_D(s) = 0$, sei $q(r|s) = 0$.

Sei $\Omega = \mathcal{S} \times \mathcal{S}$ der Stichprobenraum. Die Auswahl der Einheiten $p : \Omega \rightarrow [0, 1]$ ist durch $p(s, r) = q(r|s)p_D(s)$ definiert. Die Menge der Ereignisse sei $\mathcal{H} = \{A \subseteq \Omega\}$ und mit Hilfe von p lassen sich Wahrscheinlichkeiten für Ereignisse $A \in \mathcal{H}$ berechnen. Für das Ereignis⁷ $\{k \in s\}$, d.h. die Einheit $k \in U$ ist durch das Stichprobendesign ausgewählt worden, ist die Wahrscheinlichkeit⁸

$$\pi_k = \sum_{\{k \in s\}} p(s, r) = \sum_{\{k \in s\}} q(r|s)p_D(s) = \sum_{\{s \ni k\}} p_D(s).$$

Analog ist für das Ereignis $\{k, l \in s\}$ die Wahrscheinlichkeit

$$\pi_{kl} = \sum_{\{k, l \in s\}} p(s, r) = \sum_{\{k, l \in s\}} q(r|s)p_D(s) = \sum_{\{s \ni k, l\}} p_D(s).$$

Sei $Z : \Omega \rightarrow \mathbb{R}^d$ eine messbare Zufallsgröße⁹ und $\mathcal{F} \subseteq \mathcal{H}$ eine Mengen-Algebra bzw. σ -Algebra. Der bedingte Erwartungswert $E_p[Z|\mathcal{F}] : \Omega \rightarrow \mathbb{R}^d$ erfüllt

$$E_p[1_B E_p[Z|\mathcal{F}]] = E_p[1_B Z] \tag{2.6}$$

für jedes $B \in \mathcal{F}$. Der bedingte Erwartungswert ist nur fast sicher eindeutig, d.h. bis auf Ereignisse $B \in \mathcal{F}$ mit $P(B) = \sum_B p(s, r) = 0$. Mit dem Konzept des bedingten Erwartungswertes sind die bedingten Wahrscheinlichkeiten eng verknüpft. Die bedingte Wahrscheinlichkeit für ein Ereignis $A \in \mathcal{H}$ ist $P(A|\mathcal{F}) = E_p[1_A|\mathcal{F}]$. Betrachtet man $\mathcal{F} = \{\{\}, \Omega, B, B^c\}$ für $B \in \mathcal{H}$, dann gilt $P(A|\mathcal{F})(s, r)P(B) = P(A \cap B)$ für $(s, r) \in B$. Dies folgt aus der Definitionsgleichung des bedingten Erwartungswertes und aus der Tatsache, dass $E_p[1_A|\mathcal{F}]$ konstant auf B ist. Damit lässt sich die bedingte Wahrscheinlichkeit für das Ereignis A gegeben B schreiben als $P(A|B)$.

Sei nun $\mathcal{F} = \{A \times \mathcal{S} : A \subseteq \mathcal{S}\}$ die verfügbare Information nach Auswahl der ersten Phase. Zum Beispiel ist das Ereignis $\{k \in s\}$ in \mathcal{F} enthalten. Das Ereignis $\{k \in r\}$ ist jedoch nicht enthalten. Es ist $\mathcal{F} \subseteq \mathcal{H}$ und \mathcal{F} ist eine Mengen-Algebra. Eine Version des bedingten Erwartungswertes $E_p[Z|\mathcal{F}] : \Omega \rightarrow \mathbb{R}^d$ ist durch $E_p[Z|\mathcal{F}](s) = \sum_{\mathcal{S}} Z(s, r)q(r|s)$ gegeben, da $E_p[Z|\mathcal{F}]$ konstant auf $B = \{s\} \times \mathcal{S} \in \mathcal{F}$

⁷ $\{k \in s\}$ ist die Kurzform von $\{(s, r) \in \Omega : k \in s\}$ und $\{k, l \in s\}$ ist die Kurzform von $\{(s, r) \in \Omega : k, l \in s\}$.

⁸ $\{(s, r) \in \Omega : k \in s\}$ und $\{(s, r) \in \Omega : k, l \in s\}$ sind Ereignisse der Form $A \times \mathcal{S}$ wobei $A \subseteq \mathcal{S}$ ist. Im ersten Fall ist $A = \{s \ni k\} = \{s \in \mathcal{S} : k \in s\}$ und im zweiten Fall ist $A = \{s \ni k, l\} = \{s \in \mathcal{S} : k, l \in s\}$.

⁹ \mathbb{R}^d ist im Folgenden der d -dimensionale euklidische Raum mit $\mathbb{R}^1 = \mathbb{R}$. Er ist ausgestattet mit der borelschen σ -Algebra $\mathcal{B}(\mathbb{R}^d)$. $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ ist sowohl ein topologischer als auch ein messbarer Raum. Eine Funktion $Z : \Omega \rightarrow \mathbb{R}^d$ ist \mathcal{H} -messbar, falls $\{Z(s, r) \in A\}$ Element von \mathcal{H} für jedes $A \in \mathcal{B}(\mathbb{R}^d)$ ist.

ist. Definiert man

$$\theta_k(s) = \begin{cases} \sum_{\mathcal{S}} 1_{\{k \in r\}}(s, r) q(r|s) & k \in s \text{ und } p_D(s) > 0 \\ 1 & \text{sonst} \end{cases}$$

dann ist θ_k keine Version des bedingten Erwartungswertes $E_p[1_{\{k \in r\}}|\mathcal{F}]$. Denn 2.6 muss auch für $B = \{k \notin s\} \in \mathcal{F}$ gelten. Jedoch ist $E_p[1_B \theta_k] \neq E_p[1_B 1_{\{k \in r\}}]$. Aber es gilt $E_p[1_{\{k \in r\}}|\mathcal{F}] = \theta_k 1_{\{k \in s\}}$ fast sicher. $\theta_k: \Omega \rightarrow \mathbb{R}$ ist die Response-Wahrscheinlichkeit für die Einheit $k \in U$. Analog definiert man

$$\theta_{kl}(s) = \begin{cases} \sum_{\mathcal{S}} 1_{\{k, l \in r\}}(s, r) q(r|s) & k, l \in s \text{ und } p_D(s) > 0 \\ 1 & \text{sonst} \end{cases}$$

und es gilt $E_p[1_{\{k, l \in r\}}|\mathcal{F}] = \theta_{kl} 1_{\{k, l \in s\}}$ fast sicher.

Sei nun $w_{Mk}: \Omega \rightarrow \mathbb{R}$ mit $w_{Mk} = (\pi_k \theta_k)^{-1}$ für $k \in U$. Bei Quasi-Randomization ist

$$\hat{Y}_M = \frac{1}{N} \sum_U 1_{\{k \in r\}} w_{Mk} y_k \quad (2.7)$$

ein erwartungstreu Schätzer. Denn unter dem Stichprobendesign und dem Response-Prozeß gilt:

$$\begin{aligned} E_p[\hat{Y}_M] &= E_p[E_p[\hat{Y}_M|\mathcal{F}]] = \frac{1}{N} \sum_U E_p[E_p[1_{\{k \in r\}}|\mathcal{F}] w_{Mk}] y_k \\ &= \frac{1}{N} \sum_U E_p[1_{\{k \in s\}} \theta_k w_{Mk}] y_k \\ &= \frac{1}{N} \sum_U E_p[1_{\{k \in s\}}] w_{Mk} y_k = \frac{1}{N} \sum_U y_k \end{aligned}$$

da die Zufallsgrößen w_{Mk} für jedes $k \in U$ \mathcal{F} -messbar sind. Man beachte, dass $E_p[\hat{Y}_M|\mathcal{F}]$ (fast sicher) der Horvitz-Thompson Schätzer unter dem Design p_D ist. Man beachte auch, dass die Wahrscheinlichkeit für die Auswahl der Einheit $k \in U$ und, dass das Untersuchungsmerkmal beobachtbar ist, $P(\{k \in r\}) = E_p[1_{\{k \in s\}} \theta_k]$ gilt. Die erwartete Stichprobengröße ist $n = \sum_U E_p[1_{\{k \in s\}} \theta_k]$.

Mit Hilfe des bedingten Erwartungswertes lässt sich die bedingte Varianz definieren: $V_p[Z|\mathcal{F}] = E_p[(Z - E_p[Z|\mathcal{F}])^2|\mathcal{F}] = E_p[Z^2|\mathcal{F}] - E_p[Z|\mathcal{F}]^2$ wobei $Z: \Omega \rightarrow \mathbb{R}$ eine Zufallsgröße ist. Die Gesamtvarianz von Z lässt sich dann in zwei Teile zerlegen: $V_p[Z] = V_p[E_p[Z|\mathcal{F}]] + E_p[V_p[Z|\mathcal{F}]]$. Der erste Teil ist durch die Varianz des Horvitz-Thompson Schätzers unter dem Design p_D gegeben. Für den zweiten Teil beachte man, dass

$$E_p[1_{\{k, l \in r\}} w_{Mk} w_{Ml}|\mathcal{F}] = E_p[1_{\{k, l \in r\}}|\mathcal{F}] w_{Mk} w_{Ml} = 1_{\{k, l \in s\}} \theta_{kl} w_{Mk} w_{Ml}$$

gilt. Folglich ist die Varianz von \hat{Y}_M :

$$V_p[\hat{Y}_M] = \frac{1}{N^2} \sum \sum_U \Delta_{kl} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} + \frac{1}{N^2} E_p \left[\sum \sum_U 1_{\{k, l \in s\}} \Delta_{Mkl} (w_{Mk} y_k) (w_{Ml} y_l) \right] \quad (2.8)$$

mit $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$ und $\Delta_{Mkl} = \theta_{kl} - \theta_k \theta_l$. Die Varianz lässt sich erwartungstreu schätzen durch

$$\hat{V}_M = \frac{1}{N^2} \sum \sum_U 1_{\{k, l \in r\}} \frac{\Delta_{kl}}{\pi_{kl} \theta_{kl}} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} + \frac{1}{N^2} \sum \sum_U 1_{\{k, l \in r\}} \frac{\Delta_{Mkl}}{\theta_{kl}} (w_{Mk} y_k) (w_{Ml} y_l) \quad (2.9)$$

Wie man an (2.8) sieht, ist die Varianz bei Nonresponse größer.

Die Wahrscheinlichkeiten θ_k , $k \in U$, sind unbekannt und daher kann 2.7 nicht direkt angewendet werden. Man kann auch nicht auf die Response-Wahrscheinlichkeiten verzichten, da der Schätzer $N^{-1} \sum_U 1_{\{k \in r\}} w_k y_k$ verzerrt ist. Folglich müssen die Response-Wahrscheinlichkeiten ersetzt werden.

Sei $f_M : A \times B \rightarrow [0, 1]$ eine Funktion wobei $A \subset \mathbb{R}^d$ und $B \subset \mathbb{R}^d$. A ist der Parameter-Raum und B ist der Prädiktor-Raum, wobei die lineare Hülle von B die Dimension d hat. Im weiteren Verlauf der Arbeit werden nur Verfahren, bei denen $1/\theta_k$ durch $f_M(\hat{\lambda}, \mathbf{x}_k)$ ersetzt wird, betrachtet. Hierbei ist $\hat{\lambda} : \Omega \rightarrow A$ eine Zufallsgröße und $\mathbf{x}_k \in B$ sind individuelle Merkmalsausprägungen, die nach der Stichprobenziehung für alle Einheiten $k \in s$ beobachtbar sind. Das Ersetzen von w_{Mk} durch $\hat{w}_{Mk} = w_k f_M(\hat{\lambda}, \mathbf{x}_k)$ in 2.7 ergibt

$$\hat{Y}_M = \frac{1}{N} \sum_U 1_{\{k \in r\}} \hat{w}_{Mk} y_k \quad (2.10)$$

Dies ist im Allgemeinen kein linearer Schätzer mehr. Der Erwartungswert und die Varianz eines nicht linearen Schätzers lassen sich oft nicht exakt berechnen. Jedoch lassen sich nicht lineare Schätzer durch den Satz von Taylor linearisieren. Ein approximativer Ausdruck für den Erwartungswert bzw. für die Varianz basiert dann auf dem linearen Anteil der Taylor-Entwicklung. Der Fehler der Approximation hängt von einem nicht linearen Restglied ab. Lässt man die Grundgesamtheit beliebig groß werden, lassen sich unter gewissen Annahmen die asymptotischen Konzepte aus dem frequentistischen Ansatz der Statistik auf den Fehler übertragen.

2.3 Asymptotische Betrachtungsweise

Um nicht lineare Schätzer zu vergleichen bzw. zu bewerten ist die Nutzung des Konzepts von Fuller und Isaki (1981) und Isaki und Fuller (1982) hilfreich. Dieser Ansatz geht davon aus, dass die Größe der Grundgesamtheit gegen unendlich wächst. Hierbei betrachtet man eine Folge U_t , $t \in \mathbb{N}$, von Grundgesamtheiten der Größen N_t , wobei die Grundgesamtheiten ineinander geschachtelt sind, d.h. $U_t \subset U_{t+1}$ und $N_t < N_{t+1}$. Für jede Grundgesamtheit U_t sei Ω_t der Stichprobenraum. Die Stichprobenräume sind geschachtelt: $\Omega_t \subset \Omega_{t+1}$, $t \in \mathbb{N}$. Für die Auswahl der Einheiten, $p_t : \Omega_t \rightarrow [0, 1]$, $t \in \mathbb{N}$, gilt, dass die erwarteten Stichprobengrößen anwachsen, d.h. $n_t < n_{t+1}$. Im Falle von Quasi-Randomization besteht $p_t : \Omega_t \rightarrow [0, 1]$, $t \in \mathbb{N}$, aus zwei Komponenten, dem Stichprobendesign $p_{Dt} : \mathcal{S}_t \rightarrow [0, 1]$ und dem Response-Prozess $q_t(\cdot|s) : \mathcal{S}_t \rightarrow [0, 1]$ mit $s \in \mathcal{S}_t$, wobei \mathcal{S}_t die Menge der möglichen Stichproben aus U_t ist.

Um Folgen von Zufallsgrößen zu betrachten sei $\Omega = \prod_{t \in \mathbb{N}} \Omega_t$ der Produktraum. Im Falle von Quasi-Randomization ist Ω die Menge aller Abbildungen $(s, r) : \mathbb{N} \rightarrow \Omega_\infty$ mit $(s, r)_t \in \Omega_t$ und $\Omega_\infty = \bigcup_{t \in \mathbb{N}} \Omega_t$. Außerdem sei \mathcal{H} die kleinste σ -Algebra, so dass alle natürlichen Projektionen $\text{proj}_t : \Omega \rightarrow \Omega_t$ mit $\text{proj}_t(s, r) = (s, r)_t$, $t \in \mathbb{N}$, messbare Abbildungen sind. $\mathcal{H}_t = \{\text{proj}_t^{-1}(A) : A \subseteq \Omega_t\}$ sind alle möglichen Ereignisse bei der Auswahl aus der Grundgesamtheit U_t . Mit Kolmogorovs Erweiterungssatz¹⁰ gibt es ein eindeutiges Wahrscheinlichkeitsmaß P auf (Ω, \mathcal{H}) , so dass für jedes $t \in \mathbb{N}$ und $A_t \subseteq \Omega_t$,

$$P(\{(s, r)_t \in A_t\}) = \sum_{A_t} p_t[(s, r)_t]$$

gilt.

¹⁰Mit der diskreten Metrik sind alle Teilmengen von Ω_t offen. Da es nur endlich viele Teilmengen gibt, ist Ω_t ein kompakter metrischer Raum, wobei die Potenzmenge von Ω_t eine Borelsche σ -Algebra ist.

Bei dieser Betrachtungsweise sind die Ereignisse $\{k \in s_t\}$, $\{k, l \in s_t\}$, $\{k \in r_t\}$ und $\{k, l \in r_t\}$ für $k, l \in U_t$ und alle $t \in \mathbb{N}$, über die natürlichen Projektionen definiert und in \mathcal{H}_t enthalten. Beispielsweise ist $\{k \in s_t\} = \{\text{proj}_t^{-1}(s, r)_t : (s, r)_t \in \Omega_t \text{ und } k \in s_t\}$. Die zugehörigen Indikatorfunktionen $1_A : \Omega \rightarrow \mathbb{R}$ für Ereignisse $A \in \mathcal{H}_t$ sind \mathcal{H}_t -messbar. Obwohl die Indikatorfunktionen für Folgen aus Ω definiert sind, ändern sie ihren Wert nur, falls sich $(s, r)_t \in \Omega_t$ ändert. Sie können also auch als Funktionen $1_A : \Omega_t \rightarrow \mathbb{R}$, $A \in \mathcal{H}_t$, betrachtet werden. Die durch das Stichprobendesign induzierten Auswahlwahrscheinlichkeiten sind

$$\pi_{kt} = \mathbb{E} [1_{\{k \in s_t\}}] \quad \text{und} \quad \pi_{klt} = \mathbb{E} [1_{\{k, l \in s_t\}}]$$

für $k, l \in U_t$ und alle $t \in \mathbb{N}$. Sei außerdem $\mathcal{F}_t = \{\text{proj}_t^{-1}(A \times \mathcal{S}_t) : A \subseteq \mathcal{S}_t\}$ die verfügbare Information nach Auswahl mit dem Stichprobendesign aus der Grundgesamtheit U_t . Die Response-Wahrscheinlichkeiten $\theta_{kt} : \Omega_\infty \rightarrow [0, 1]$ und $\theta_{klt} : \Omega_\infty \rightarrow [0, 1]$ können wie im letzten Abschnitt für $k, l \in U_t$ und jedes $t \in \mathbb{N}$ definiert werden, bzw. sind implizit durch die Gleichungen

$$\mathbb{E} [1_{\{k \in s_t\}} | \mathcal{F}_t] = 1_{\{k \in s_t\}} \theta_{kt} \quad \text{und} \quad \mathbb{E} [1_{\{k, l \in s_t\}} | \mathcal{F}_t] = 1_{\{k, l \in s_t\}} \theta_{klt}$$

fast sicher eindeutig gegeben.

Mit dem obigen Konzept und dem einhergehenden Wahrscheinlichkeitsraum (Ω, \mathcal{H}, P) lassen sich für Folgen von Schätzfunktionen $\hat{Y}_t : \Omega \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, definiert durch

$$\hat{Y}_t = \frac{1}{N_t} \sum_{U_t} 1_{\{k \in r_t\}} w_{Mkt} y_k \quad \text{mit} \quad w_{Mkt} = \frac{1}{\pi_{kt} \theta_{kt}} \quad \text{für } k \in U_t, t \in \mathbb{N}$$

asymptotische Betrachtungsweisen anstellen. Man beachte, dass ohne Nonresponse unter \hat{Y}_t der Horvitz-Thompson Schätzer zu verstehen ist.

Um den Schätzer (2.10) bewerten zu können, müssen zunächst Annahmen bzgl. der asymptotischen Eigenschaften seines theoretischen Pendant (2.7) getroffen werden. Hierbei geht es also um die Frage, wie sich \hat{Y}_t zu $Y_t = \mathbb{E}[\hat{Y}_t]$ asymptotisch verhält. Der Abstand zwischen zwei Zufallsgrößen kann durch verschiedene Metriken ausgedrückt werden. \hat{Y}_t ist dann für große $t \in \mathbb{N}$ bezüglich der gewählten Metrik nahe an Y_t . Hierzu wird das Konzept der Schwachen Konvergenz, der Konvergenz in Wahrscheinlichkeit und der gleichgradige Integrierbarkeit benötigt. Um diese Konzepte zu erläutern, wird die Maß- und Integrationstheorie vorausgesetzt. Die folgenden Ausführungen geben eine Einführung in diese Konzepte.

Sei $Z_t : \Omega \rightarrow \mathbb{R}^d$, $t \in \mathbb{N}$, eine Folge von Zufallsgrößen wie zum Beispiel $\sqrt{n_t}(\hat{Y}_t - Y_t)$. Bei schwacher Konvergenz betrachtet man nicht die Zufallsgröße direkt, sondern die durch sie induzierten Wahrscheinlichkeitsmaße $P_t = P \circ Z_t^{-1} : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $t \in \mathbb{N}$:

Definition (Schwache Konvergenz). Die Folge von Wahrscheinlichkeitsmaßen $P_t : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $t \in \mathbb{N}$, konvergiert schwach gegen $P_\infty : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, falls für alle stetigen und beschränkten Funktionen $f : \mathbb{R}^d \rightarrow \mathbb{R}$

$$\lim_{t \rightarrow \infty} \int f dP_t = \int f dP_\infty$$

gilt.

Diese Definition basiert allein auf einer Folge von Wahrscheinlichkeitsmaßen. Falls $P_t = P \circ Z_t^{-1}$ schwach gegen $P_\infty : \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ konvergiert, dann heißt die Folge von Zufallsgrößen Z_t , $t \in \mathbb{N}$, schwach konvergent. Wegen

$$\mathbb{E}[f(Z_t)] = \int f(Z_t) dP = \int f dP_t$$

gilt $\lim_{t \rightarrow \infty} \mathbb{E}[f(Z_t)] = \mathbb{E}[f(Z_\infty)]$ für alle stetigen und beschränkten Funktionen $f: \mathbb{R}^d \rightarrow \mathbb{R}$ und jede Zufallsgröße $Z_\infty: \Omega \rightarrow \mathbb{R}^d$ mit $P_\infty = P \circ Z_\infty^{-1}$.

Um den Begriff der schwachen Konvergenz nutzen zu können, sind weitere äquivalente Darstellungen geeignet. Diese Darstellungen sind im Rahmen des Portmanteau-Theorems zusammengefaßt.

Satz (Portmanteau-Theorem). Sei $P_t: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $t \in \mathbb{N}$, eine Folge von Wahrscheinlichkeitsmaßen. Die folgenden Aussagen sind äquivalent.

1. $\lim_{t \rightarrow \infty} \int f dP_t = \int f dP_\infty$ für alle stetigen und beschränkten bzw. für alle beschränkten Lipschitz-stetigen Funktionen $f: \mathbb{R}^d \rightarrow \mathbb{R}$
2. $\limsup_{t \rightarrow \infty} \int 1_A dP_t \leq \int 1_A dP_\infty$ für alle abgeschlossenen Mengen $A \subset \mathbb{R}^d$
3. $\liminf_{t \rightarrow \infty} \int 1_A dP_t \geq \int 1_A dP_\infty$ für alle offenen Mengen $A \subset \mathbb{R}^d$
4. $\lim_{t \rightarrow \infty} \int 1_A dP_t = \int 1_A dP_\infty$ für alle $A \subset \mathbb{R}^d$ mit $\int 1_{\text{Cl}A \setminus \text{Int}A} dP_\infty = 0$ wobei $\text{Cl}A$ der Abschluss von A und $\text{Int}A$ das Innere von A ist
5. $\lim_{t \rightarrow \infty} \int f dP_t = \int f dP_\infty$ für alle beschränkten Funktionen $f: \mathbb{R}^d \rightarrow \mathbb{R}$, die P_∞ -fast überall stetig sind

Beweis. Sei $A \subset \mathbb{R}^d$ eine abgeschlossene Menge. Seien die gleichmäßig stetigen Funktionen $f_k: \mathbb{R}^d \rightarrow \mathbb{R}$, $k \in \mathbb{N}$, durch

$$f_k(z) = 1 - \min\{k \inf_{a \in A} \|z - a\|, 1\}$$

definiert. Diese Funktionen sind Lipschitz-stetig und konvergieren punktweise gegen die Funktion $1_A: \mathbb{R}^d \rightarrow \mathbb{R}$. Dann folgt aus der Aussage 1 und des Satzes von Lebesgue die Aussage 2:

$$\int 1_A dP_\infty = \lim_{k \rightarrow \infty} \int f_k dP_\infty = \lim_{k \rightarrow \infty} \limsup_{t \rightarrow \infty} \int f_k dP_t \geq \limsup_{t \rightarrow \infty} \int 1_A dP_t$$

Die Aussage 3 folgt aus der Aussage 2 durch Komplementbildung. Aus beiden Aussagen zusammen folgt

$$\begin{aligned} \int 1_{\text{Int}A} dP_\infty &\leq \liminf_{t \rightarrow \infty} \int 1_{\text{Int}A} dP_t \leq \liminf_{t \rightarrow \infty} \int 1_A dP_t \\ &\leq \limsup_{t \rightarrow \infty} \int 1_A dP_t \leq \limsup_{t \rightarrow \infty} \int 1_{\text{Cl}A} dP_t \leq \int 1_{\text{Cl}A} dP_\infty \end{aligned}$$

Da der Rand nach Voraussetzung das Maß Null hat, folgt Aussage 4. Ohne Einschränkung sei nun $0 \leq f \leq 1$ für die P_∞ -fast überall stetig Funktion $f: \mathbb{R}^d \rightarrow \mathbb{R}$. Mit der Aussage 4 konvergiert die

Folge

$$P_t(\{f \geq a\}) = \int 1_{\{f \geq a\}} dP_t, \quad t \in \mathbb{N}$$

für alle $a \in \mathbb{R}$, wobei der Rand von $\{f \geq a\}$ das Maß Null hat, gegen

$$P_\infty(\{f \geq a\}) = \int 1_{\{f \geq a\}} dP_\infty$$

Es gibt aber nur höchstens abzählbar viele $a \in \mathbb{R}$, für die der Rand von $\{f \geq a\}$ ein Maß größer als Null hat. Denn der Rand ist Teilmenge von $B \cup \{f = a\}$ wobei B die Menge der Unstetigkeitsstellen von f ist, welche nach Voraussetzung das P_∞ -Maß Null hat und

$$P_\infty(\{f = a\}) = \int 1_{\{f = a\}} dP_\infty$$

kann nur für höchstens abzählbar viele $a \in \mathbb{R}$ einen Wert größer als Null haben.¹¹ Also konvergiert $P_t(\{f \geq a\})$ als Funktion von $a \in \mathbb{R}$ bezüglich des Lebesgue-Maßes fast überall gegen $P_\infty(\{f \geq a\})$. Dann folgt mit dem Satz von Lebesgue die Aussage 5:

$$\int f dP_\infty = \int_0^1 P_\infty(\{f \geq a\}) da = \lim_{t \rightarrow \infty} \int_0^1 P_t(\{f \geq a\}) da = \lim_{t \rightarrow \infty} \int f dP_t$$

Aussage 1 folgt trivialerweise aus Aussage 5. □

Mit der Aussage 5 des Portmanteau-Theorems folgt das sogenannte Continuous Mapping Theorem. Dieses Theorem macht eine Aussage darüber, dass stetige Funktionen angewendet auf Zufallsgrößen deren Eigenschaft der schwachen Konvergenz beibehalten.

Korollar (Continuous Mapping Theorem). Seien $Z_t: \Omega \rightarrow \mathbb{R}^d$, $t \in \mathbb{N}$, und $Z_\infty: \Omega \rightarrow \mathbb{R}^d$ Zufallsgrößen mit der Eigenschaft, dass $P_t = P \circ Z_t^{-1}: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ schwach gegen $P_\infty: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ konvergiert. Außerdem sei $f: \mathbb{R}^d \rightarrow \mathbb{R}^{d'}$ eine P_∞ -fast überall stetige Funktion. Dann konvergiert die Folge $f \circ Z_t$, $t \in \mathbb{N}$, schwach gegen $f \circ Z_\infty$.

Sei $Z_t: \Omega \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, eine Folge von Zufallsgrößen, die schwach gegen $Z_\infty: \Omega \rightarrow \mathbb{R}$ konvergiert. Mit dem Continuous Mapping Theorem konvergiert auch $|Z_t|$, $t \in \mathbb{N}$, schwach gegen $|Z_\infty|$. Außerdem gibt es nur höchstens abzählbar viele Stellen $a \in \mathbb{R}$ mit $P(\{|Z_\infty| = a\}) > 0$. Also gibt es zu jedem $\varepsilon > 0$ ein $M_0 > 0$, so dass $P(\{|Z_\infty| > M_0\}) < \varepsilon/2$ und $P(\{|Z_\infty| = M_0\}) = 0$ gilt. Mit Aussage 4 des Portmanteau-Theorems konvergiert $P(\{|Z_t| > M_0\})$ gegen $P(\{|Z_\infty| > M_0\})$. Es gibt also ein $t_0 \in \mathbb{N}$, so dass

$$P(\{|Z_t| > M_0\}) \leq P(\{|Z_\infty| > M_0\}) + \left| P(\{|Z_t| > M_0\}) - P(\{|Z_\infty| > M_0\}) \right| < \varepsilon$$

für alle $t > t_0$ ist. Für die endlich vielen $t \leq t_0$ gibt es ein $M_1 > 0$, so dass $P(\{|Z_t| > M_1\}) < \varepsilon$ ist. Damit ist $P(\{|Z_t| > M\}) < \varepsilon$ für $M = \max\{M_0, M_1\}$ und alle $t \in \mathbb{N}$. Der mehrdimensionale Fall lässt sich analog zeigen. Eine schwach konvergente Folge von Zufallsgrößen ist also in Wahrscheinlichkeit beschränkt.

¹¹ Es kann höchstens $k \in \mathbb{N}$ Stellen $a \in \mathbb{R}$ geben mit $P_\infty(\{f = a\}) \geq 1/k$. Sonst wäre es kein Wahrscheinlichkeitsmaß. Die Vereinigung über $k \in \mathbb{N}$ ist eine abzählbare Vereinigung endlicher Stellen und somit abzählbar.

Definition (Beschränktheit in Wahrscheinlichkeit). $Z_t : \Omega \rightarrow \mathbb{R}^d$, $t \in \mathbb{N}$, ist in Wahrscheinlichkeit beschränkt, falls es für jedes $\varepsilon > 0$ ein $M > 0$ existiert, so dass $P(\{\|Z_t\| > M\}) < \varepsilon$ für alle $t \in \mathbb{N}$ gilt.

Die Schreibweise $Z_t = O_p(1)$ bedeutet, dass die Folge Z_t , $t \in \mathbb{N}$, in Wahrscheinlichkeit beschränkt ist. Falls es eine Folge reeller Zahlen $a_t > 0$, $t \in \mathbb{N}$, gibt, so dass $a_t^{-1}Z_t$ in Wahrscheinlichkeit beschränkt ist, dann wird die Schreibweise $Z_t = O_p(a_t)$ benutzt. Die Folge a_t , $t \in \mathbb{N}$, stellt eine asymptotisch obere Schranke dar. Falls $Z_t : \Omega \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, und $\tilde{Z}_t : \Omega \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, in Wahrscheinlichkeit beschränkt sind, dann ist auch das Produkt $Z_t\tilde{Z}_t$ in Wahrscheinlichkeit beschränkt.¹² Folglich gilt $Z_t\tilde{Z}_t = O_p(a_t\tilde{a}_t)$, falls $Z_t = O_p(a_t)$ und $\tilde{Z}_t = O_p(\tilde{a}_t)$ gilt. Die Summe $Z_t + \tilde{Z}_t$ ist ebenfalls in Wahrscheinlichkeit beschränkt¹³ und es gilt $Z_t + \tilde{Z}_t = O_p(\max\{a_t, \tilde{a}_t\})$, falls $Z_t = O_p(a_t)$ und $\tilde{Z}_t = O_p(\tilde{a}_t)$ gilt.

Sei $Z_t = (Z_{1t}, \dots, Z_{dt}) : \Omega \rightarrow \mathbb{R}^d$, $t \in \mathbb{N}$, und $Z_{it} = O_p(a_{it})$ für $i = 1, \dots, d$. Dann ist $Z_t = O_p(\max_i\{a_{it}\})$, da $\|Z_t\| \leq \sum_i |Z_{it}|$ gilt. D.h. die asymptotische obere Schranke eines Vektors oder einer Matrix wird durch die maximale obere Schranke der Einzelkomponenten bestimmt. Sei $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ eine P_∞ -fast überall Lipschitz-stetige Funktion und $\hat{Y}_t - Y_t = O_p(a_t)$ wobei $\hat{Y}_t : \Omega \rightarrow \mathbb{R}^d$ eine Folge von Schätzfunktionen für $Y_t \in \mathbb{R}^d$ ist. Dann ist auch $f(\hat{Y}_t) - f(Y_t) = O_p(a_t)$.

Falls $\sqrt{n_t}(\hat{Y}_t - Y_t)$, $t \in \mathbb{N}$, schwach konvergiert, dann ist die Folge in Wahrscheinlichkeit beschränkt. Zu vorgegebenen $\varepsilon > 0$ gibt es ein $M > 0$, so dass $P(\{\sqrt{n_t}|\hat{Y}_t - Y_t| > M\}) < \varepsilon$ für alle $t \in \mathbb{N}$ gilt. Außerdem gibt es zu $\delta > 0$ ein $t_0 \in \mathbb{N}$ mit $\delta\sqrt{n_t} > M$. Somit ist $P(\{|\hat{Y}_t - Y_t| > \delta\}) < \varepsilon$ für $t > t_0$. Also konvergiert $\hat{Y}_t - Y_t$ in Wahrscheinlichkeit gegen Null¹⁴:

Definition (Konvergenz in Wahrscheinlichkeit). $Z_t : \Omega \rightarrow \mathbb{R}^d$, $t \in \mathbb{N}$, konvergiert in Wahrscheinlichkeit gegen $Z_\infty : \Omega \rightarrow \mathbb{R}^d$, falls es für jedes $\varepsilon > 0$ und jedes $\delta > 0$ ein $t_0 \in \mathbb{N}$ existiert, so dass $P(\{\|Z_t - Z_\infty\| > \delta\}) < \varepsilon$ für alle $t > t_0$ gilt.

Falls $Z_t = (Z_{1t}, \dots, Z_{dt}) : \Omega \rightarrow \mathbb{R}^d$ in Wahrscheinlichkeit gegen $Z_\infty = (Z_{1\infty}, \dots, Z_{d\infty}) : \Omega \rightarrow \mathbb{R}^d$ konvergiert, dann konvergiert wegen $|Z_{it} - Z_{i\infty}| \leq \|Z_t - Z_\infty\|$ für $i = 1, \dots, d$ auch jede Komponente von Z_t . Falls $|Z_{it} - Z_{i\infty}| \leq \delta/d$ für $i = 1, \dots, d$ dann ist $\|Z_t - Z_\infty\| \leq \sum_i |Z_{it} - Z_{i\infty}| \leq \delta$ und es gilt

$$P(\{\|Z_t - Z_\infty\| > \delta\}) \leq \sum_i P(\{|Z_{it} - Z_{i\infty}| > \delta/d\})$$

Falls $Z_{it} : \Omega \rightarrow \mathbb{R}$ für jedes $i = 1, \dots, d$ in Wahrscheinlichkeit gegen $Z_{i\infty} : \Omega \rightarrow \mathbb{R}$ konvergiert, dann konvergiert auch $Z_t = (Z_{1t}, \dots, Z_{dt})$ gegen $Z_\infty = (Z_{1\infty}, \dots, Z_{d\infty})$.

Sei $\hat{Y}_t : \Omega \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, ein Folge von Schätzfunktionen für die entsprechenden Größen $Y_t \in \mathbb{R}$. Falls $Z_t = \hat{Y}_t - Y_t$ in Wahrscheinlichkeit gegen 0 konvergiert, dann heißt \hat{Y}_t Design-konsistent für Y_t . Man beachte, dass die Differenz in Design-Wahrscheinlichkeit gegen einen fixen Wert konvergiert. Falls \hat{Y}_t , $t \in \mathbb{N}$, selbst gegen einen endlichen Wert konvergieren soll, muss zusätzlich gefordert werden, dass $Y_\infty = \lim_{t \rightarrow \infty} Y_t$ endlich ist. Wegen $|\hat{Y}_t - Y_\infty| \leq |\hat{Y}_t - Y_t| + |Y_t - Y_\infty|$ konvergiert daher \hat{Y}_t in Design-Wahrscheinlichkeit gegen den fixen Wert Y_∞ .

Es besteht der folgende Zusammenhang zwischen schwacher Konvergenz und der Konvergenz in Wahrscheinlichkeit. Falls $Z_t : \Omega \rightarrow \mathbb{R}^d$, $t \in \mathbb{N}$, schwach gegen eine Konstante $a \in \mathbb{R}^d$ konvergiert,

¹² $\{|Z_t| \leq M\} \cap \{|\tilde{Z}_t| \leq \tilde{M}\}$ ist Teilmenge von $\{|Z_t\tilde{Z}_t| \leq M\tilde{M}\}$ und somit ist $P(\{|Z_t\tilde{Z}_t| > M\tilde{M}\}) \leq P(\{|Z_t| > M\}) + P(\{|\tilde{Z}_t| > \tilde{M}\})$ für beliebige $M > 0$ und $\tilde{M} > 0$.

¹³ $\{|Z_t| \leq M/2\} \cap \{|\tilde{Z}_t| \leq M/2\}$ ist Teilmenge von $\{|Z_t| + |\tilde{Z}_t| \leq M\}$ und $|Z_t + \tilde{Z}_t| \leq |Z_t| + |\tilde{Z}_t|$. Damit ist $P(\{|Z_t + \tilde{Z}_t| > M\}) \leq P(\{|Z_t| > M/2\}) + P(\{|\tilde{Z}_t| > M/2\})$ für beliebiges $M > 0$.

¹⁴Die Annahme bei Deville und Särndal (1992), dass $\hat{Y}_t - Y_t$ bei gleichzeitiger schwacher Konvergenz von $\sqrt{n_t}(\hat{Y}_t - Y_t)$ in Wahrscheinlichkeit gegen Null konvergiert, ist daher redundant.

dann konvergiert die Folge auch in Wahrscheinlichkeit gegen diese Konstante. Sei $P_\infty: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ wobei für $B \in \mathcal{B}(\mathbb{R}^d)$, $a \in B$, $P_\infty(B) = 1$ gilt. Dann hat $A(\delta) = \{z \in \mathbb{R}^d : \|z - a\| \leq \delta\}$ für jedes $\delta > 0$ einen Rand mit P_∞ -Maß Null und selbst das P_∞ -Maß Eins. Daher ist für jedes $\delta > 0$ mit dem Portmanteau-Theorem

$$\lim_{t \rightarrow \infty} P(\{\|Z_t - a\| \leq \delta\}) = \lim_{t \rightarrow \infty} \int 1_{A(\delta)} dP_t = \int 1_{A(\delta)} dP_\infty = 1$$

Umgekehrt folgt aus der Konvergenz in Wahrscheinlichkeit die schwache Konvergenz. Hierzu sei $A \subset \mathbb{R}^d$ eine abgeschlossene Menge und $A_\delta = \{z \in \mathbb{R}^d : \inf_{a \in A} \|z - a\| \leq \delta\}$ eine Menge, die um einen δ -Rand größer als A ist. Seien $Z_t, t \in \mathbb{N}$, und Z_∞ Zufallsgrößen. Mit der Definition von A_δ gilt für jedes $t \in \mathbb{N}$

$$\begin{aligned} P(\{Z_t \in A\}) &= P(\{Z_t \in A, \|Z_t - Z_\infty\| > \delta\}) + P(\{Z_t \in A, \|Z_t - Z_\infty\| \leq \delta\}) \\ &\leq P(\{\|Z_t - Z_\infty\| > \delta\}) + P(\{Z_\infty \in A_\delta\}) \end{aligned}$$

Falls $Z_t, t \in \mathbb{N}$, in Wahrscheinlichkeit gegen Z_∞ konvergiert, gilt

$$\limsup_{t \rightarrow \infty} P(\{Z_t \in A\}) \leq P(\{Z_\infty \in A_\delta\})$$

für jedes $\delta > 0$. Wegen $A = \bigcap_{\delta > 0} A_\delta$ und der monotonen Konvergenz-Eigenschaft von Maßen folgt mit Aussage 2 des Portmanteau-Theorems, dass $Z_t, t \in \mathbb{N}$, schwach gegen Z_∞ konvergiert.

Ersetzt man in dieser Argumentation Z_∞ durch eine schwach gegen Z_∞ konvergente Folge, $\tilde{Z}_t, t \in \mathbb{N}$, dann konvergiert $Z_t, t \in \mathbb{N}$, schwach gegen Z_∞ , falls die Differenz $Z_t - \tilde{Z}_t$ in Wahrscheinlichkeit gegen Null konvergiert. Dies kann man auf folgende Weise nutzen. Falls $\tilde{Z}_t, t \in \mathbb{N}$, schwach gegen eine Konstante $a \in \mathbb{R}^d$ konvergiert und $Z_t, t \in \mathbb{N}$, schwach gegen Z_∞ konvergiert, dann konvergiert (Z_t, a) schwach gegen (Z_∞, a) und (Z_t, \tilde{Z}_t) in Wahrscheinlichkeit gegen (Z_t, a) . Somit konvergiert (Z_t, \tilde{Z}_t) schwach gegen (Z_∞, a) .

Sei $\hat{Y}_t: \Omega \rightarrow \mathbb{R}$ der Horvitz-Thompson Schätzer für den Mittelwert $Y_t \in \mathbb{R}$. Falls $\hat{Y}_t - Y_t$ schwach gegen Null konvergiert und der Grenzwert $\lim_{t \rightarrow \infty} Y_t = Y_\infty$ endlich ist, dann konvergiert $(\hat{Y}_t - Y_t, Y_t)$ schwach gegen $(0, Y_\infty)$. Mit dem Continuous Mapping Theorem konvergiert dann auch (\hat{Y}_t, Y_t) gegen (Y_∞, Y_∞) . Aber auch die Differenz $f(\hat{Y}_t) - f(Y_t)$ konvergiert schwach gegen Null für jede Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$, die in einer offenen Umgebung von Y_∞ stetig ist. Da die Differenz gegen Null konvergiert, konvergiert $f(\hat{Y}_t) - f(Y_t)$ auch in Wahrscheinlichkeit gegen Null.

Aus den beiden Annahmen von Deville und Särndal (1992), der Grenzwert $Y_\infty = \lim_{t \rightarrow \infty} Y_t$ existiert für den Mittelwert in der Grundgesamtheit und $\sqrt{n_t}(\hat{Y}_t - Y_t)$ konvergiert schwach, wobei \hat{Y}_t der Horvitz-Thompson Schätzer für Y_t ist, folgt, dass $\hat{Y}_t - Y_t$ in Wahrscheinlichkeit gegen Null konvergiert. Weiterhin konvergiert \hat{Y}_t auch in Wahrscheinlichkeit gegen Y_∞ . Darauf basierend konvergiert auch die Differenz $f(\hat{Y}_t) - f(Y_t)$ in Wahrscheinlichkeit gegen Null. Ein Verhältniswert läßt sich damit auf Basis von zwei Horvitz-Thompson Schätzern Design-konsistent schätzen.

Der nächste Satz zeigt, dass stetige Abbildungen die Konvergenz in Wahrscheinlichkeit erhalten:

Satz (Continuous Mapping Theorem, Konvergenz in Wahrscheinlichkeit). Seien $Z_t: \Omega \rightarrow \mathbb{R}^d, t \in \mathbb{N}$, und $Z_\infty: \Omega \rightarrow \mathbb{R}^d$ Zufallsgrößen mit der Eigenschaft, dass die Folge $Z_t, t \in \mathbb{N}$, in Wahrscheinlichkeit gegen Z_∞ konvergiert. Außerdem sei $f: \mathbb{R}^d \rightarrow \mathbb{R}^d$ eine $P_\infty = P \circ Z_\infty^{-1}$ -fast überall stetige Funktion. Dann konvergiert die Folge $f \circ Z_t, t \in \mathbb{N}$, in Wahrscheinlichkeit gegen $f \circ Z_\infty$.

Beweis. Angenommen $f \circ Z_t, t \in \mathbb{N}$, konvergiert nicht in Wahrscheinlichkeit gegen $f \circ Z_\infty$. Dann gibt es $\varepsilon > 0, \delta > 0$ und eine Teilfolge $Z_{\varphi(t)}, t \in \mathbb{N}, \varphi : \mathbb{N} \rightarrow \mathbb{N}$ monoton, so dass $P(\{\|f(Z_{\varphi(t)}) - f(Z_\infty)\| > \varepsilon\}) > \delta$ für alle $t \in \mathbb{N}$ gilt. Nach Voraussetzung konvergiert $Z_{\varphi(t)}, t \in \mathbb{N}$, in Wahrscheinlichkeit gegen Z_∞ . Jede in Wahrscheinlichkeit konvergente Folge hat eine fast sichere konvergente Teilfolge. Daher gibt es eine weitere Teilfolge $Z_{\tilde{\varphi}(\varphi(t))}, t \in \mathbb{N}, \tilde{\varphi} : \mathbb{N} \rightarrow \varphi(\mathbb{N})$ monoton, so dass diese fast sicher gegen Z_∞ konvergiert. Mit dem Folgenkriterium für Stetigkeit konvergiert auch $f \circ Z_{\tilde{\varphi}(\varphi(t))}, t \in \mathbb{N}$, fast sicher gegen $f \circ Z_\infty$. Aber dann konvergiert diese auch in Wahrscheinlichkeit gegen $f \circ Z_\infty$. Somit ist Null Häufungspunkt der Folge $P(\{\|f(Z_{\varphi(t)}) - f(Z_\infty)\| > \varepsilon\}), t \in \mathbb{N}$, Widerspruch. \square

Eine schwach konvergente Folge von Zufallsgrößen ist in Wahrscheinlichkeit beschränkt. Sie konvergiert in Wahrscheinlichkeit, falls sie eine asymptotische obere Schranke besitzt, die gegen Null konvergiert. Hier stellt sich nun die Frage, wann $\lim_{t \rightarrow \infty} E[Z_t] = E[Z_\infty]$ gilt, falls die Folge $Z_t, t \in \mathbb{N}$, schwach gegen Z_∞ konvergiert. Denn der Grenzwert der Erwartungswerte muss nicht mit dem Erwartungswert des Grenzwertes übereinstimmen. Mit dem Lemma von Fatou und der Aussage 3 des Portmanteau-Theorems folgt

$$E[Z_\infty] = \int_0^\infty P(\{Z_\infty > a\}) da \leq \liminf_{t \rightarrow \infty} \int_0^\infty P(\{Z_t > a\}) da = \liminf_{t \rightarrow \infty} E[Z_t]$$

wobei $Z_t : \Omega \rightarrow \mathbb{R}, t \in \mathbb{N}$, eine Folge von nicht-negativen Zufallsgrößen ist, die schwach gegen $Z_\infty : \Omega \rightarrow \mathbb{R}$ konvergiert. Die schwächste Forderung, damit $\lim_{t \rightarrow \infty} E[Z_t] = E[Z_\infty]$ bei schwacher Konvergenz gilt, ist, dass die Folge $Z_t, t \in \mathbb{N}$, gleichgradig integrierbar ist. Das wird im Folgenden gezeigt.

Definition (Gleichgradige Integrierbarkeit). Eine Folge $Z_t : \Omega \rightarrow \mathbb{R}^d, t \in \mathbb{N}$, heißt gleichgradig integrierbar, falls es für jedes $\varepsilon > 0$ ein $M > 0$ existiert, so dass

$$E[\|Z_t\|1_{\{\|Z_t\| > M\}}] < \varepsilon$$

für alle $t \in \mathbb{N}$ gilt.

Falls Z_t gleichgradig integrierbar ist, sind wegen $|Z_{it}| \leq \|Z_t\|$ für $i = 1, \dots, d$ auch alle Einzelkomponenten gleichgradig integrierbar:

$$E[|Z_{it}|1_{\{|Z_{it}| > M\}}] \leq E[\|Z_t\|1_{\{\|Z_t\| > M\}}]$$

Umgekehrt ist bei gleichgradiger Integrierbarkeit der Einzelkomponenten wegen $\|Z_t\| \leq d \max\{Z_{it}\}$ auch Z_t gleichgradig integrierbar:

$$E[\|Z_t\|1_{\{\|Z_t\| > M\}}] \leq \sum_{i=1}^d E[d|Z_{it}|1_{\{d|Z_{it}| > M\}}]$$

Deshalb gelten die folgenden Aussagen auch für den mehrdimensionalen Fall. Man beachte, dass aus der Beschränktheit von $E[\|Z_t\|]$ nicht die gleichgradige Integrierbarkeit folgt. Jedoch folgt die Beschränktheit von $E[\|Z_t\|]$ aus der gleichgradigen Integrierbarkeit. Aber die Folge $Z_t, t \in \mathbb{N}$, ist gleichgradig integrierbar, falls $E[\|Z_t\|^a], a > 1$, beschränkt ist, da

$$E[|Z_t|1_{\{|Z_t| > M\}}] \leq M^{1-a} E[\|Z_t\|^a]$$

gilt.

Die nächsten beiden Sätze zeigen, dass bei schwacher Konvergenz die gleichgradige Integrierbarkeit äquivalent zur Konvergenz im Mittel ist, Billingsley (1999).

Satz. Die Folge $Z_t : \Omega \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, sei gleichgradig integrierbar und konvergiere schwach gegen $Z_\infty : \Omega \rightarrow \mathbb{R}$. Dann ist Z_∞ integrierbar und es gilt $\lim_{t \rightarrow \infty} E[Z_t] = E[Z_\infty]$.

Beweis. Für eine gleichgradig integrierbare Folge gilt $E[|Z_t|] < \infty$, $t \in \mathbb{N}$. Also ist $\liminf_{t \rightarrow \infty} E[Z_t]$ endlich und mit der Aussage 3 des Portmanteau-Theorems Z_∞ integrierbar, siehe oben. Sei Z_t^+ der Positivteil und Z_t^- der Negativteil von Z_t . Mit dem Continuous Mapping Theorem konvergiert Z_t^+ , $t \in \mathbb{N}$, schwach gegen Z_∞^+ und Z_t^- , $t \in \mathbb{N}$, schwach gegen Z_∞^- . Wegen $Z_t = Z_t^+ - Z_t^-$ reicht es aus, die obige Behauptung für eine nicht negative Folge zu zeigen. Sei $\varepsilon > 0$. Da die Folge gleichgradig integrierbar ist gibt es ein $M_1 > 0$ mit $E[Z_t 1_{\{Z_t > M_1\}}] < \varepsilon/3$. Da der Erwartungswert absolut stetig ist, gibt es ein $M_2 > 0$ mit $E[Z_\infty 1_{\{Z_\infty > M_2\}}] < \varepsilon/3$. Wählt man nun $M > \max\{M_1, M_2\}$ mit $P(\{Z_\infty = M\}) = 0$, dann folgt mit der Aussage 4 des Portmanteau-Theorems $P(\{a < Z_t \leq M\}) \rightarrow P(\{a < Z_\infty \leq M\})$ für jedes $a < M$ mit $P(\{Z_\infty = a\}) = 0$. Da es nur abzählbar viele Stellen $a < M$ mit $P(\{Z_\infty = a\}) > 0$ gibt, folgt mit dem Satz von Lebesgue

$$\lim_{t \rightarrow \infty} \int_0^M P(\{a < Z_t \leq M\}) da = \int_0^M P(\{a < Z_\infty \leq M\}) da$$

Somit gibt es ein $t_0 \in \mathbb{N}$, so dass die absolute Differenz der Integrale kleiner $\varepsilon/3$ für alle $t > t_0$ ist. Also gibt es für jedes $\varepsilon > 0$ ein $t_0 \in \mathbb{N}$ mit

$$\left| E[Z_t] - E[Z_\infty] \right| \leq \left| E[Z_t 1_{\{Z_t \leq M\}}] - E[Z_\infty 1_{\{Z_\infty \leq M\}}] \right| + E[Z_t 1_{\{Z_t > M\}}] + E[Z_\infty 1_{\{Z_\infty > M\}}] < \varepsilon$$

für alle $t > t_0$. □

Sei $\hat{Y}_t : \Omega \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, eine Folge von Schätzfunktionen für die entsprechenden Größen in der Grundgesamtheit $Y_t \in \mathbb{R}$. Falls $n_t(\hat{Y}_t - Y_t)^2$ gleichgradig integrierbar ist und zusätzlich $\sqrt{n_t}(\hat{Y}_t - Y_t)$ schwach konvergent ist, dann konvergiert mit dem Continuous Mapping Theorem $n_t(\hat{Y}_t - Y_t)^2$ schwach gegen eine Zufallsgröße Z_∞ und mit dem obigen Satz folgt, dass der Grenzwert

$$\lim_{t \rightarrow \infty} E[n_t(\hat{Y}_t - Y_t)^2] = E[Z_\infty]$$

endlich ist. Insbesondere konvergiert der Mittlere Quadratische Fehler $E[(\hat{Y}_t - Y_t)^2]$ gegen Null. Seine asymptotische obere Schranke ist $O(n_t^{-1})$. Falls \hat{Y}_t der Horvitz-Thompson Schätzer für den Mittelwert in der Grundgesamtheit ist, dann konvergiert die Varianz des Horvitz-Thompson Schätzers gegen Null.

Der folgende Satz zeigt, dass die gleichgradig Integrierbarkeit auch ein notwendiges Kriterium bei schwacher Konvergenz ist, falls der Grenzwert $\lim_{t \rightarrow \infty} E[n_t(\hat{Y}_t - Y_t)^2]$ existiert.

Satz. Seien $Z_t : \Omega \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, und $Z_\infty : \Omega \rightarrow \mathbb{R}$ nicht-negativ und integrierbar und die Folge Z_t , $t \in \mathbb{N}$, konvergiert schwach gegen Z_∞ . Dann ist die Folge Z_t , $t \in \mathbb{N}$, gleichgradig integrierbar, falls $\lim_{t \rightarrow \infty} E[Z_t] = E[Z_\infty]$ gilt.

Beweis. Nach Voraussetzung und dem Portmanteau-Theorem ist

$$\lim_{t \rightarrow \infty} E[Z_t 1_{\{Z_t > M\}}] = E[Z_\infty 1_{\{Z_\infty > M\}}]$$

für jedes $M > 0$ mit $P(\{Z_t = M\}) = 0$. Wie bei der Beschränktheit in Wahrscheinlichkeit können M und $t_0 \in \mathbb{N}$ so gewählt werden, dass

$$\mathbb{E} [Z_t 1_{\{Z_t > M\}}] < \mathbb{E} [Z_\infty 1_{\{Z_\infty > M\}}] + \left| \mathbb{E} [Z_t 1_{\{Z_t > M\}}] - \mathbb{E} [Z_\infty 1_{\{Z_\infty > M\}}] \right| < \varepsilon$$

für alle $t > t_0$ ist. Berücksichtigt man in M noch alle $t \leq t_0$, dann folgt die Behauptung. \square

Sei \hat{Y}_t der Horvitz-Thompson Schätzer für den Mittelwert Y_t in der Grundgesamtheit. Eine Annahme bei Deville und Särndal (1992) besagt, dass $\sqrt{n_t}(\hat{Y}_t - Y_t)$ gegen eine Zufallsgröße Z_∞ mit Erwartungswert $\mathbb{E}[Z_\infty] = 0$ und einer fixen Varianz $V[Z_\infty]$ konvergiert. Sie folgern daraus, dass der Grenzwert von $\mathbb{E}[n_t(\hat{Y}_t - Y_t)^2]$ mit der Varianz $V[Z_\infty]$ übereinstimmt. Das setzt aber voraus, dass die Folge $n_t(\hat{Y}_t - Y_t)^2$, $t \in \mathbb{N}$, gleichgradig integrierbar ist.¹⁵¹⁶

Falls \hat{Y}_t der Horvitz-Thompson Schätzer für Y_t ist und der Mittlere Quadratische Fehler ($\mathbb{E}[(\hat{Y}_t - Y_t)^2]$) $O(n_t^{-1})$ ist, dann gilt mit der Markov-Ungleichung $P(\{\sqrt{n_t}|\hat{Y}_t - Y_t| > M\}) < \varepsilon$ für beliebiges $\varepsilon > 0$, wobei

$$M = \frac{(\sup_{t \in \mathbb{N}} \mathbb{E}[n_t(\hat{Y}_t - Y_t)^2])^{1/2}}{\varepsilon}$$

ist. Also ist $\sqrt{n_t}(\hat{Y}_t - Y_t) = O_p(1)$ und $\hat{Y}_t - Y_t$ konvergiert in Wahrscheinlichkeit gegen Null. $(\hat{Y}_t - Y_t)^2$ ist daher mit dem obigen Satz gleichgradig integrierbar.

Fuller und Isaki (1981) geben für den Fall einer geschichteten Stichprobe mit Ziehung einer Einheit je Schicht hinreichende Bedingungen für die Design-Konsistenz des Horvitz-Thompson Schätzers an. Dies lässt sich auf den Mikrozensus übertragen. Sei n_t die Anzahl der Schichten, welche der Anzahl der zufällig ausgewählten Einheiten entspricht. M_E sei die Anzahl der Einheiten je Schicht. Insgesamt gibt es $N_t = n_t M_E$ Einheiten. Es ist $\pi_{kt} = M_E^{-1}$ und $\pi_{klt} = 0$, falls k und l in der selben Schicht sind und $\pi_{klt} = M_E^{-2}$, falls k und l in unterschiedlichen Schichten sind. y_k ist der Wert der Untersuchungsvariable der k -ten Einheit. Falls $y_k < M_B$ für $k = 1, 2, \dots$ gilt - das ist zum Beispiel für dichotome Merkmale der Fall - ist mit 2.4 die Varianz des Horvitz-Thompson Schätzer

$$V[\hat{Y}_t] = \frac{1}{2N_t^2} \sum \sum_{U_t} (\pi_{kt}\pi_{lt} - \pi_{klt}) \left(\frac{y_k}{\pi_{kt}} - \frac{y_l}{\pi_{lt}} \right)^2 \leq n_t^{-1} M_E^{-1} (M_E - 1) (2M_B)^2$$

und somit gilt für den Horvitz-Thompson Schätzer $\sqrt{n_t}(\hat{Y}_t - Y_t) = O_p(1)$.

Unter Weglassung des Index t , sei die Gültigkeit der folgenden Annahmen für \hat{Y}_M definiert in (2.7) und $Y = \mathbb{E}[\hat{Y}_M]$ vorausgesetzt:

[A1] Der Grenzwert $\lim Y$ existiert.

[A2] $\sqrt{n}(\hat{Y}_M - Y)$ konvergiert schwach.

[A3] $n^p(\hat{Y}_M - Y)^{2p}$ ist gleichgradig integrierbar für ein $p \geq 1$.

¹⁵Falls $\sqrt{n_t}(\hat{Y}_t - Y_t)$ selbst normalverteilt wäre, dann ist dessen Charakteristische Funktion $\varphi_t(a) = \exp\{-a^2 \mathbb{E}[n_t(\hat{Y}_t - Y_t)^2]/2\}$. Diese konvergiert für jedes a gegen die Charakteristische Funktion $\varphi_\infty(a) = \exp\{-a^2 V[Z_\infty]/2\}$. Also ist $\lim_{t \rightarrow \infty} \mathbb{E}[n_t(\hat{Y}_t - Y_t)^2] = V[Z_\infty]$ und daher ist $n_t(\hat{Y}_t - Y_t)^2$ gleichgradig integrierbar.

¹⁶Die Annahme, dass $\lim_{t \rightarrow \infty} Y_t$ endlich ist, kann nicht auf die Varianz angewendet werden, da für die Design-Gewichte w_k , $k \in \mathbb{N}$, die Bestandteil der Varianz sind, keine Einschränkungen vorgenommen werden.

Aus der Annahme [A3] folgt, dass $n(\hat{Y}_M - Y)^2$ gleichgradig integrierbar ist. Mit Annahme [A2] existiert der Grenzwert von $E[n(\hat{Y}_M - Y)^2]$. Das bedeutet, dass die Varianz $O(n^{-1})$ ist. Mit der Jensenschen Ungleichung ist

$$E[E[\hat{Y}_M - Y|\mathcal{F}]^2] \leq E[(\hat{Y}_M - Y)^2]$$

Da $E[\hat{Y}_M - Y|\mathcal{F}]$ fast sicher $\hat{Y}_{HT} - Y$ ist, ist die Varianz von $\hat{Y}_{HT} - Y$ auch $O(n^{-1})$.

Es lässt sich zeigen, dass mit der Annahme [A3] und $p = 8$ das Restglied bei der Anwendung des Satzes von Taylor im quadratischen Mittel mit Rate $O(n^{-2})$ gegen Null konvergiert. Die Annahmen [A1], [A2] und [A3] beziehen sich auf den Schätzer (2.7) und nicht auf den Schätzer (2.10). Der Schätzer (2.10) ist abhängig von einem geschätzten Parameter. Der nächste Abschnitt behandelt parameterabhängige Zufallsgrößen und wie die Annahmen [A1], [A2] und [A3] auf diese angewendet werden können.

2.4 Gleichmäßige Konvergenz

In diesem Abschnitt wird eine Folge von parameterabhängigen Zufallsgrößen $\hat{X}_t: \Omega \times A \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, wie zum Beispiel

$$\hat{X}_t(\lambda) = \frac{1}{N_t} \sum_{U_t} 1_{\{k \in r_t\}} w_{Mkt} f(\lambda, \mathbf{x}_k) \quad (2.11)$$

betrachtet. Die Funktion $f: A \times B \rightarrow \mathbb{R}$, wobei $A \subset \mathbb{R}^d$ und $B \subset \mathbb{R}^d$, ist stetig und $\mathbf{x}_k \in B$ sind individuelle Charakteristiken, die die Untersuchungsvariable y enthalten dürfen. Die Funktionen \hat{X}_t sind fast sicher stetig und schwanken zufällig um die Funktionen $X_t: A \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, definiert durch

$$X_t(\lambda) = \frac{1}{N_t} \sum_{U_t} f(\lambda, \mathbf{x}_k) \quad (2.12)$$

Ersetzt man in (2.11) λ durch einen Schätzer $\hat{\lambda}_t: \Omega \rightarrow A$ und in (2.12) λ durch einen festen Wert $\lambda_0 \in A$, dann stellt sich die Frage, welche Annahmen erfüllt sein müssen, damit $\hat{X}_t(\hat{\lambda}_t) - X_t(\lambda_0)$ in Wahrscheinlichkeit bzw. im p -ten Mittel gegen Null konvergiert unter der Bedingung, dass $\hat{\lambda}_t$ in Wahrscheinlichkeit gegen λ_0 konvergiert. Ein Beispiel hierfür ist (2.10), indem f durch $f(\lambda, \mathbf{x}_k) = y_k f_M(\hat{\lambda}, \mathbf{x}_k) / f_M(\lambda_0, \mathbf{x}_k)$ ersetzt wird.

Dieses Problem lässt sich in zwei Teilprobleme zerlegen, $\hat{X}_t(\hat{\lambda}_t) - X_t(\hat{\lambda}_t)$ und $X_t(\hat{\lambda}_t) - X_t(\lambda_0)$. Für beide Terme muß die Konvergenz in Wahrscheinlichkeit gezeigt werden. Für den ersten Term wäre es vorteilhaft, wenn $\hat{X}_t - X_t$ gleichmäßig gegen die Nullfunktion konvergieren würde. In diesem Fall wäre die Folge $\hat{\lambda}_t$, $t \in \mathbb{N}$, von untergeordneter Bedeutung. Nach Voraussetzung ist $\hat{X}_t - X_t$ fast sicher stetig. Allerdings garantiert die fast sichere Stetigkeit der einzelnen Folgenglieder nicht, dass bei punktwiser Anwendung der Annahme [A2] die Grenzfunktion der Nullfunktion entspricht und damit stetig ist. Damit die Grenzfunktion stetig ist, muss die Folge gewissermaßen stochastisch gleichmäßig gegen die Grenzfunktion konvergieren.

Definition (Stochastisch Gleichmäßige Konvergenz). $Z_t: \Omega \times A \rightarrow \mathbb{R}$, $A \subset \mathbb{R}^d$, $t \in \mathbb{N}$, konvergiert stochastisch gleichmäßig gegen $Z_\infty: \Omega \times A \rightarrow \mathbb{R}$, $A \subset \mathbb{R}^d$, falls es für jedes $\varepsilon > 0$ und jedes $\delta > 0$ ein $t_0 \in \mathbb{N}$ existiert, so dass $P(\{\|Z_t - Z_\infty\|_A > \delta\}) < \varepsilon$ für alle $t > t_0$ gilt. Hierbei ist $\|\cdot\|_A$ die Supremumsnorm auf der Menge A .

Falls die Differenz $\hat{X}_t - X_t$ nicht nur punktweise in Wahrscheinlichkeit gegen Null konvergiert, wären für große $t \in \mathbb{N}$ die parameterabhängige Zufallsgröße \hat{X}_t und die Funktionen X_t in diesem Sinne nicht unterscheidbar. Anders ausgedrückt, gilt für jedes $\delta > 0$

$$\lim_{t \rightarrow \infty} P(\{\|\hat{X}_t - X_t\|_A > \delta\}) = 0$$

Im Folgenden wird gezeigt, dass bei punktwiser Gültigkeit der Annahmen [A1] und [A2] sowie einer bestimmten Regularitätsbedingung an f , die stochastisch gleichmäßige Konvergenz folgt. Mit der zusätzlichen Annahme [A3] und der Kompaktheit des Parameter-Raums folgt sogar die Konvergenz im p -ten Mittel bezüglich der Supremumsnorm.

Eine stochastisch gleichmäßige Konvergenz setzt voraus, dass $Z_\infty : \Omega \times A \rightarrow \mathbb{R}$ existiert. Bei Vollständigkeit des Raums in dem die Folge betrachtet wird, ist es ausreichend Cauchy-Folgen zu betrachten.

Definition (Stochastisch Cauchy). $Z_t : \Omega \times A \rightarrow \mathbb{R}$, $A \subset \mathbb{R}^d$, $t \in \mathbb{N}$, ist eine stochastische Cauchy-Folge, falls es für jedes $\varepsilon > 0$ und jedes $\delta > 0$ ein $t_0 \in \mathbb{N}$ existiert, so dass $P(\{\|Z_i - Z_j\|_A > \delta\}) < \varepsilon$ für alle $i, j > t_0$ gilt.

Der Raum, der auf einer kompakten Menge $A \subset \mathbb{R}^d$ stetigen Funktionen, der mit der Supremumsnorm ausgestattet ist, ist ein Banach-Raum und ist somit vollständig. Dieser wird im Folgenden mit $C(A)$ bezeichnet. In einem vollständigen Raum konvergiert jede Cauchy-Folge. Der nächste Satz zeigt, dass der Grenzwert für stochastische Cauchy-Folgen von stetigen parameterabhängigen Zufallsgrößen in $C(A)$ existiert, siehe zum Beispiel Kallenberg (2001).

Satz (Vollständigkeit). Sei $Z_t : \Omega \times A \rightarrow \mathbb{R}$, A kompakt, $t \in \mathbb{N}$, eine stochastische Cauchy-Folge von parameterabhängigen Zufallsgrößen in $C(A)$. Dann existiert eine parameterabhängige Zufallsgröße $Z_\infty : \Omega \times A \rightarrow \mathbb{R}$ in $C(A)$, so dass Z_t , $t \in \mathbb{N}$, stochastisch gleichmäßig gegen Z_∞ konvergiert.

Beweis. Nach Voraussetzung gibt es eine monotone Abbildung $\varphi : \mathbb{N} \rightarrow \mathbb{N}$, so dass

$$P(\{\|Z_{\varphi(t+1)} - Z_{\varphi(t)}\|_A > 2^{-t}\}) < 2^{-t}$$

für alle $t \in \mathbb{N}$ gilt. Mit dem Borel-Cantelli Lemma folgt, dass $P(\{\|Z_{\varphi(t+1)} - Z_{\varphi(t)}\|_A > 2^{-t}$ unendlich oft $\}) = 0$ ist und somit $\sum_{t=1}^{\infty} \|Z_{\varphi(t+1)} - Z_{\varphi(t)}\|_A$ fast sicher endlich ist. Für ein beliebiges $\varepsilon > 0$ existiert also ein $i \in \mathbb{N}$, so dass

$$\|Z_{\varphi(j)} - Z_{\varphi(i)}\|_A \leq \sum_{t=i}^j \|Z_{\varphi(t+1)} - Z_{\varphi(t)}\|_A \leq \sum_{t=i}^{\infty} \|Z_{\varphi(t+1)} - Z_{\varphi(t)}\|_A < \varepsilon$$

für alle $j > i$ ist. Damit ist $Z_{\varphi(t)}$, $t \in \mathbb{N}$, fast sicher eine Cauchy-Folge in $C(A)$. Definiert man $Z_\infty(s, r) = \lim_{t \rightarrow \infty} Z_t(s, r)$ für alle $(s, r) \in \Omega$, falls die Grenzfunktion existiert und $Z_\infty(s, r) = 0$ sonst. Die so definierten Grenzfunktionen sind stetig, da $C(A)$ vollständig ist und $Z_{\varphi(t)}$, $t \in \mathbb{N}$, fast sicher gegen $Z_\infty : \Omega \times A \rightarrow \mathbb{R}$ konvergiert. Eine fast sicher konvergente Folge, konvergiert auch in Wahrscheinlichkeit. Also konvergieren beide Terme auf der rechten Seite der Ungleichung

$$P(\{\|Z_t - Z_\infty\|_A > \delta\}) \leq P(\{\|Z_t - Z_{\varphi(t)}\|_A > \delta/2\}) + P(\{\|Z_{\varphi(t)} - Z_\infty\|_A > \delta/2\})$$

gegen Null. □

Man beachte, dass $Z_\infty \subset C(A)$ fast sicher beschränkt ist, da Bilder von kompakten Mengen unter stetigen Abbildungen wieder kompakt sind. Somit ist $P(\bigcup_{i \in \mathbb{N}} \{\|Z_\infty\|_A \leq i\}) = 1$. Mit der monotonen Konvergenzeigenschaft von Maßen gibt es zu $\varepsilon > 0$ ein $M > 0$, so dass $P(\{\|Z_\infty\|_A > M\}) < \varepsilon$ ist. Wegen

$$\|Z_t\|_A \leq \|Z_\infty\|_A + \|Z_t - Z_\infty\|_A$$

ist die Folge $Z_t, t \in \mathbb{N}$, in $C(A)$ in Wahrscheinlichkeit beschränkt.

Sei der Erwartungswert von $\|Z_t\|_A^p, p \geq 1$, für alle $t \in \mathbb{N}$ endlich. Dann ist $E[\|Z_t\|_A^p]$ für jedes $t \in \mathbb{N}$ eine stetige Funktion¹⁷ in $C(A)$. Existiert für jedes $\varepsilon > 0$ ein $t \in \mathbb{N}$, so dass

$$E[\|Z_i - Z_j\|_A^p] \leq \varepsilon$$

für alle $i, j > t$ gilt, dann ist die Folge $Z_t, t \in \mathbb{N}$, eine Cauchy-Folge von Zufallsgrößen in $C(A)$ mit endlichem Erwartungswert von $\|Z_t\|_A^p$. Mit der Markov-Ungleichung folgt, dass $Z_t, t \in \mathbb{N}$, auch eine stochastische Cauchy-Folge ist. Der Beweis des obigen Satzes zeigt dann, dass es eine monotone Abbildung $\varphi : \mathbb{N} \rightarrow \mathbb{N}$ gibt, so dass $Z_{\varphi(t)}, t \in \mathbb{N}$, fast sicher gegen Z_∞ in $C(A)$ konvergiert. Somit gibt es ein $t_0 \in \mathbb{N}$, so dass

$$\|Z_t - Z_{\varphi(t)}\|_A^p \leq \varepsilon$$

für alle $t > t_0$ gilt. Mit dem Lemma von Fatou ist

$$E[\|Z_t - Z_\infty\|_A^p] \leq \liminf_{i \rightarrow \infty} E[\|Z_t - Z_{\varphi(i)}\|_A^p] \leq \varepsilon$$

Der Erwartungswert von $\|Z_\infty\|_A^p$ ist damit endlich¹⁸ und $E[\|Z_\infty\|_A^p]$ ist eine stetige Funktion.

Das Arzelà-Ascoli Theorem besagt, dass, falls eine gleichgradig stetige Funktionenfolge¹⁹ punktweise konvergiert, dann konvergiert sie auf einer kompakten Menge A gleichmäßig. Mit der Folge, dass die punktweise (auf A) definierte Grenzfunktion stetig ist. Bei einer stochastischen Variante würde die gleichmäßige stochastischen Konvergenz von $\hat{X}_t - X_t$ aus der Annahme der punktweisen Konvergenz, [A2], folgen, falls $\hat{X}_t - X_t$ stochastisch gleichgradig stetig ist:

Definition (Stochastisch Gleichgradige Stetigkeit). $Z_t : \Omega \times A \rightarrow \mathbb{R}, A \subset \mathbb{R}^d$ ist stochastisch gleichgradig stetig, falls es für jedes $\eta > 0$ und $\varepsilon > 0$ ein $\delta > 0$ existiert, so dass

$$P\left(\left\{\sup_{\tilde{\lambda} \in A} \sup_{A(\tilde{\lambda}, \delta)} |Z_t(\lambda) - Z_t(\tilde{\lambda})| > \eta\right\}\right) < \varepsilon \quad (2.13)$$

für alle $t \in \mathbb{N}$ wobei $A(\tilde{\lambda}, \delta) = \{\lambda \in A : \|\lambda - \tilde{\lambda}\| < \delta\}$ eine offene Umgebungen mit einem Radius $\delta > 0$ um $\tilde{\lambda} \in A$ ist.

Der folgende Satz ist eine stochastische Variante des Arzelà-Ascoli Theorems, Newey (1991).

¹⁷ A ist eine kompakte Menge. Daher ist die Stetigkeit in jedem Punkt äquivalent zur gleichmäßigen Stetigkeit. Für eine Folge $\lambda_i, i \in \mathbb{N}$, mit $\lim_{i \rightarrow \infty} \lambda_i = \lambda_\infty$ gilt $\lim_{i \rightarrow \infty} |Z_t(s, r)(\lambda_i)|^p = |Z_t(s, r)(\lambda_\infty)|^p$ für jedes $(s, r) \in \Omega$. $|Z_t(s, r)(\lambda_i)|^p$ ist außerdem durch $\|Z_t(s, r)\|_A^p$ für jedes $(s, r) \in \Omega$ beschränkt. Mit dem Satz von Lebesgue ist dann $\lim_{i \rightarrow \infty} E[|Z_t(\lambda_i)|^p] = E[|Z_t(\lambda_\infty)|^p]$, da der Erwartungswert von $\|Z_t\|_A^p$ endlich ist.

¹⁸Vgl. auch den Satz von Riesz-Fischer für die Vollständigkeit von L^p -Räumen.

¹⁹Eine Funktionenfolge $X_t, t \in \mathbb{N}$, ist gleichmäßig gleichgradig stetig, falls es für jedes $\eta > 0$ ein $\delta > 0$ existiert, so dass für alle $\lambda, \tilde{\lambda} \in A$ und alle $t \in \mathbb{N}$: $\|\lambda - \tilde{\lambda}\| < \delta \Rightarrow |X_t(\lambda) - X_t(\tilde{\lambda})| < \eta$.

Satz (Arzelà-Ascoli). Die Folge von parameterabhängigen Zufallsgrößen $Z_t : \Omega \times A \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, in $C(A)$ konvergiert genau dann stochastisch gleichmäßige gegen $Z_\infty : \Omega \times A \rightarrow \mathbb{R}$ in $C(A)$ wenn Z_t , $t \in \mathbb{N}$, stochastisch gleichgradig stetig ist und punktweise auf A gegen Z_∞ konvergiert.

Beweis. Seien $\varepsilon > 0$ und $\eta > 0$ vorgegeben. Die Folge Z_t , $t \in \mathbb{N}$, ist stochastisch gleichgradig stetig. Daher existiert ein $t_s \in \mathbb{N}$ und ein $\delta > 0$, so dass 2.13 gilt. Da A kompakt ist, gibt es zu diesem δ endlich viele Punkte $\lambda_1, \lambda_2, \dots, \lambda_{L(\delta)}$, so dass A Teilmenge der Vereinigung der offenen Umgebungen $A(\lambda_l, \delta)$, $l = 1, 2, \dots, L(\delta)$, ist. Mit diesen Stützstellen ist

$$\|Z_i - Z_j\|_A = \max_l \sup_{A(\lambda_l, \delta)} |Z_i(\lambda) - Z_j(\lambda)|$$

und daher

$$\|Z_i - Z_j\|_A \leq \max_l \left\{ \sup_{A(\lambda_l, \delta)} |Z_i(\lambda) - Z_i(\lambda_l)| + |Z_i(\lambda_l) - Z_j(\lambda_l)| + \sup_{A(\lambda_l, \delta)} |Z_j(\lambda_l) - Z_j(\lambda)| \right\}$$

Für den zweiten Term auf der rechten Seite der Ungleichung lässt sich die punktweisen Konvergenz in Wahrscheinlichkeit anwenden. Wegen der punktweisen Konvergenz gibt es ein $t_p \in \mathbb{N}$, so dass $P(\{|Z_i(\lambda_l) - Z_j(\lambda_l)| > \eta/3\}) < \varepsilon/3L(\delta)$ für alle $l = 1, 2, \dots, L(\delta)$ und $t > t_p$ gilt. Da das Ereignis $\{\max_l |Z_i(\lambda_l) - Z_j(\lambda_l)| > \eta/3\}$ äquivalent zu dem Ereignis $\bigcup_l \{|Z_i(\lambda_l) - Z_j(\lambda_l)| > \eta/3\}$ ist, gilt

$$P(\{\max_l |Z_i(\lambda_l) - Z_j(\lambda_l)| > \eta/3\}) \leq \sum_l P(\{|Z_i(\lambda_l) - Z_j(\lambda_l)| > \eta/3\}) < \varepsilon/3$$

für alle $t > t_p$ und somit auch für alle $t > t_0 = \max\{t_p, t_s\}$. Für den ersten und den dritten Term ist die Wahrscheinlichkeit für das Ereignis $\{\max_l \sup_{A(\lambda_l, \delta)} |Z_i(\lambda) - Z_j(\lambda_l)| > \eta/3\}$ kleiner als $\varepsilon/3$ für alle $t > t_0$. Folglich gibt es für $\eta > 0$ und $\varepsilon > 0$ ein $t_0 \in \mathbb{N}$, so dass Z_t , $t \in \mathbb{N}$, eine stochastische Cauchy-Folge in $C(A)$ ist.

Sei nun die Folge Z_t , $t \in \mathbb{N}$, stochastisch gleichmäßig konvergent. Für $\varepsilon > 0$ gibt es dann ein $M > 0$, so dass $P(\{\|Z_t\|_A > M\}) < \varepsilon/2$ für alle $t \in \mathbb{N}$ ist. Die Menge $C_M = \{f \in C(A) : \|f\|_A \leq M\}$ ist kompakt. Wählt man $\eta > 0$, dann gibt es endlich viele Funktionen $f_l \in C(A)$, $l = 1, 2, \dots, L(\eta)$, so dass

$$C_M \subset \bigcap_l \{f \in C(A) : \|f - f_l\|_A \leq \eta/3\}$$

ist. Mit diesem Netz aus stetigen Funktionen gilt für alle $t \in \mathbb{N}$

$$P(\{\min_l \{\|Z_t - f_l\|_A > \eta/3\}\}) = P(\bigcap_l \{\|Z_t - f_l\|_A > \eta/3\}) \leq P(\{\|Z_t\|_A > M\}) < \varepsilon/2$$

Außerdem gibt es ein $\delta > 0$, so dass $|f_l(\lambda) - f_l(\tilde{\lambda})| < \eta/3$ für alle $\lambda, \tilde{\lambda} \in A$ mit $\|\lambda - \tilde{\lambda}\| < \delta$ und $l = 1, 2, \dots, L(\eta)$ ist. Mit der Dreiecksungleichung folgt

$$|Z_t(\lambda) - Z_t(\tilde{\lambda})| \leq 2 \min_l \|Z_t - f_l\|_A + \eta/3$$

für alle $\lambda, \tilde{\lambda} \in A$ mit $\|\lambda - \tilde{\lambda}\| < \delta$ und alle $t \in \mathbb{N}$. Da $P(\{2 \min_l \|Z_t - f_l\|_A > 2\eta/3\}) < \varepsilon$ ist, gilt 2.13 für das gewählte $\varepsilon > 0$ und das gewählte $\eta > 0$ für alle $t \in \mathbb{N}$. \square

Für Zufallsgrößen in $C(A)$ deren Supremum auf A im p -ten Mittel endlich ist, lässt sich auch eine gleichgradige Stetigkeit im p -ten Mittel definieren. Die Folge von parameterabhängigen Zufallsgrößen

$Z_t : \Omega \times A \rightarrow \mathbb{R}$, $A \subset \mathbb{R}^d$ ist im p -ten Mittel gleichgradig stetig, falls es für jedes $\eta > 0$ ein $\delta > 0$ existiert, so dass

$$\mathbb{E} \left[\sup_{\tilde{\lambda} \in A} \sup_{A(\tilde{\lambda}, \delta)} |Z_t(\lambda) - Z_t(\tilde{\lambda})|^p \right] < \eta \quad (2.14)$$

für alle $t \in \mathbb{N}$ gilt. Wegen der Ungleichung

$$\begin{aligned} \mathbb{E} [\|Z_i - Z_j\|_A^p] &\leq 3^{p-1} \left\{ \mathbb{E} \left[\max_l \sup_{A(\lambda_l, \delta)} |Z_i(\lambda) - Z_i(\lambda_l)|^p \right] + \mathbb{E} [\sum_l |Z_i(\lambda_l) - Z_j(\lambda_l)|^p] \right. \\ &\quad \left. + \mathbb{E} \left[\max_l \sup_{A(\lambda_l, \delta)} |Z_j(\lambda_l) - Z_j(\lambda)|^p \right] \right\} \end{aligned}$$

ist die Folge Z_t , $t \in \mathbb{N}$, eine Cauchy-Folge von Zufallsgrößen in $C(A)$, wenn die Folge im p -ten Mittel gleichgradig stetig ist und sie punktweise in p -ten Mittel konvergiert (Annahmen [A2] und [A3]).

Sei $\hat{\lambda}_t : \Omega \rightarrow A$, $t \in \mathbb{N}$, eine Folge von Schätzfunktionen, die in Wahrscheinlichkeit gegen $\lambda_0 \in A$ konvergiert. Sei \hat{X}_t , $t \in \mathbb{N}$, stochastisch gleichgradig stetig und X_t , $t \in \mathbb{N}$, gleichgradig stetig. Da X_t , $t \in \mathbb{N}$, gleichgradig stetig ist, gibt es für $\eta > 0$ ein $\delta > 0$, so dass für alle $t \in \mathbb{N}$ und $\lambda \in A$ mit $\|\lambda - \lambda_0\| < \delta$ auch $|X_t(\lambda) - X_t(\lambda_0)| < \eta$ gilt. Somit ist

$$\begin{aligned} P(\{|X_t(\hat{\lambda}_t) - X_t(\lambda_0)| > \eta\}) &\leq P(\{|X_t(\hat{\lambda}_t) - X_t(\lambda_0)| > \eta, \|\hat{\lambda}_t - \lambda_0\| < \delta\}) + \\ &\quad P(\{|X_t(\hat{\lambda}_t) - X_t(\lambda_0)| > \eta, \|\hat{\lambda}_t - \lambda_0\| \geq \delta\}) \\ &\leq P(\{\|\hat{\lambda}_t - \lambda_0\| \geq \delta\}) \end{aligned}$$

Also konvergiert $X_t(\hat{\lambda}_t)$ in Wahrscheinlichkeit gegen $X_t(\lambda_0)$. Außerdem konvergiert $\hat{X}_t - X_t$, $t \in \mathbb{N}$, mit Annahme [A2] und der stochastischen Variante des Satzes von Arzelà-Ascoli gleichmäßig auf jeder kompakten Menge in A , die λ_0 enthält. Die Wahrscheinlichkeit, dass die kompakte Menge $\hat{\lambda}$ nicht enthält, konvergiert gegen Null. Mit der Dreiecksungleichung

$$|\hat{X}_t(\hat{\lambda}_t) - X_t(\lambda_0)| \leq |\hat{X}_t(\hat{\lambda}_t) - X_t(\hat{\lambda}_t)| + |X_t(\hat{\lambda}_t) - X_t(\lambda_0)|$$

folgt dann, dass $\hat{X}_t(\hat{\lambda}_t)$ in Wahrscheinlichkeit gegen $X_t(\lambda_0)$ konvergiert.

Sei nun A kompakt und \hat{X}_t , $t \in \mathbb{N}$, ist im p -ten Mittel gleichgradig stetig und X_t , $t \in \mathbb{N}$, ist gleichgradig stetig. Falls X_t , $t \in \mathbb{N}$, punktweise auf der kompakten Menge A beschränkt ist (Annahme [A1]), folgt mit dem Satz von Arzelà-Ascoli, dass X_t , $t \in \mathbb{N}$, dann auch gleichmäßig beschränkt ist. D.h. es gibt insbesondere ein $M > 0$, so dass $|X_t(\lambda)|^p < M$ für alle $t \in \mathbb{N}$ und alle $\lambda \in A$ ist. Wegen der gleichgradigen Stetigkeit und der Konvergenz in Wahrscheinlichkeit von $\hat{\lambda}_t$, $t \in \mathbb{N}$, gegen λ_0 , gibt es für $\eta > 0$ ein $\delta > 0$ und ein $t_0 \in \mathbb{N}$, so dass für alle $t > t_0$ und $\lambda \in A$ mit $\|\lambda - \lambda_0\| < \delta$ sowohl $|X_t(\lambda) - X_t(\lambda_0)| < \eta/2$ als auch $P(\{\|\hat{\lambda}_t - \lambda_0\| \geq \delta\}) < \eta/8M$ gilt. Somit ist

$$\begin{aligned} \mathbb{E}[|X_t(\hat{\lambda}_t) - X_t(\lambda_0)|^p] &\leq \mathbb{E} \left[|X_t(\hat{\lambda}_t) - X_t(\lambda_0)|^p \mathbf{1}_{\{\|\hat{\lambda}_t - \lambda_0\| < \delta\}} \right] + \\ &\quad \mathbb{E} \left[|X_t(\hat{\lambda}_t) - X_t(\lambda_0)|^p \mathbf{1}_{\{\|\hat{\lambda}_t - \lambda_0\| \geq \delta\}} \right] \\ &\leq \mathbb{E} \left[|X_t(\hat{\lambda}_t) - X_t(\lambda_0)|^p \mathbf{1}_{\{\|\hat{\lambda}_t - \lambda_0\| < \delta\}} \right] + 4MP(\{\|\hat{\lambda}_t - \lambda_0\| \geq \delta\}) \end{aligned}$$

kleiner als η für alle $t > t_0$. Also konvergiert $X_t(\hat{\lambda}_t)$ im p -ten Mittel gegen $X_t(\lambda_0)$. Mit der obigen Dreiecksungleichung konvergiert $\hat{X}_t(\hat{\lambda}_t) - X_t(\lambda_0)$ im p -ten Mittel gegen Null, da $\mathbb{E}[\|\hat{X}_t - X_t\|_A^p]$, $t \in \mathbb{N}$,

mit den Annahme [A2] und [A3] gegen Null konvergiert.

Basierend auf einer Idee von Pötscher und Prucha (1989) zeigt der nächste Satz, dass unter der Annahme [A1] und den genannten Voraussetzungen für die Funktion f die stochastisch gleichgradige Stetigkeit von \hat{X}_t und die gleichgradige Stetigkeit von X_t folgt.

Satz. Sei $f: A \times B \rightarrow \mathbb{R}$, $A \subset \mathbb{R}^d$ kompakt, $B \subset \mathbb{R}^d$, eine stetige Funktion. Die Folge von parameter-abhängigen Zufallsgrößen $\hat{X}_t: \Omega \times A \rightarrow \mathbb{R}$, $t \in \mathbb{N}$, definiert durch

$$\hat{X}_t(\lambda) = \frac{1}{N_t} \sum_{U_t} 1_{\{k \in r_t\}} w_{Mkt} f(\lambda, \mathbf{x}_k)$$

ist stochastisch gleichgradig stetig, falls es eine Funktion $f_s: B \rightarrow \mathbb{R}^+$ mit den folgenden beiden Eigenschaften gibt:

1. $\|f(\cdot, \mathbf{x})\|_A \leq f_s(\mathbf{x})$ für alle $\mathbf{x} \in B$
2. $\sup N_t^{-1} \sum_{U_t} \|\mathbf{x}_k\| f_s(\mathbf{x}_k) < \infty$

Beweis. Zuerst definiert man eine monoton steigende Folge kompakter Menge wie folgt: $B_j = \{\mathbf{x} \in B : \|\mathbf{x}\| \leq j\}$, $j \in \mathbb{N}$. Die Vereinigung der Menge überdeckt B . Es gibt zwei Möglichkeiten. Entweder ist $\mathbf{x}_k \in B_j$ oder $\mathbf{x}_k \notin B_j$. Die Grundgesamtheit lässt sich damit aufteilen in $U_{tj} = \{k \in U_t : \mathbf{x}_k \in B_j\}$ und $\bar{U}_{tj} = \{k \in U_t : \mathbf{x}_k \notin B_j\}$. Mit den Definitionen von U_{tj} und \bar{U}_{tj} ist

$$\begin{aligned} |\hat{X}_t(\lambda) - \hat{X}_t(\tilde{\lambda})| &\leq N_t^{-1} \sum_{U_t} 1_{\{k \in r\}} w_{Mkt} |f(\lambda, \mathbf{x}_k) - f(\tilde{\lambda}, \mathbf{x}_k)| \\ &\leq N_t^{-1} \sum_{U_{tj}} 1_{\{k \in r\}} w_{Mkt} |f(\lambda, \mathbf{x}_k) - f(\tilde{\lambda}, \mathbf{x}_k)| + \\ &\quad N_t^{-1} \sum_{\bar{U}_{tj}} 1_{\{k \in r\}} w_{Mkt} 2 \|f(\cdot, \mathbf{x}_k)\|_A \end{aligned}$$

Seien außerdem die Funktionen $h_j: A \times A \rightarrow \mathbb{R}$, $j \in \mathbb{N}$, für jedes Paar $(\lambda, \tilde{\lambda}) \in A \times A$ durch

$$h_j(\lambda, \tilde{\lambda}) = \|f(\lambda, \cdot) - f(\tilde{\lambda}, \cdot)\|_{B_j}$$

definiert. Da die Funktion f stetig ist, gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$, so dass $|f(\lambda, \mathbf{x}) - f(\tilde{\lambda}, \tilde{\mathbf{x}})| < \varepsilon$ für alle $(\lambda, \mathbf{x}), (\tilde{\lambda}, \tilde{\mathbf{x}}) \in A \times B$ mit $\|(\lambda, \mathbf{x}) - (\tilde{\lambda}, \tilde{\mathbf{x}})\| < \delta$ ist. Dann ist aber auch $|f(\lambda, \mathbf{x}) - f(\tilde{\lambda}, \mathbf{x})| < \varepsilon$ für alle $\mathbf{x} \in B_j$ und alle $\lambda, \tilde{\lambda} \in A$ mit $\|\lambda - \tilde{\lambda}\| < \delta$. Somit gibt es zu jedem $\varepsilon > 0$ und $j \in \mathbb{N}$ ein $\delta > 0$, so dass $h_j(\lambda, \tilde{\lambda}) < \varepsilon$ für alle $\lambda, \tilde{\lambda} \in A$ mit $\|\lambda - \tilde{\lambda}\| < \delta$ gilt. Da $|f(\lambda, \mathbf{x}_k) - f(\tilde{\lambda}, \mathbf{x}_k)| \leq h_j(\lambda, \tilde{\lambda})$ für alle $k \in U_{tj}$ und $\|z_k\|/j > 1$ für alle $k \in \bar{U}_{tj}$ gilt, ist

$$|\hat{X}_t(\lambda) - \hat{X}_t(\tilde{\lambda})| \leq h_j(\lambda, \tilde{\lambda}) N_t^{-1} \sum_{U_{tj}} 1_{\{k \in r\}} w_{Mkt} + 2(N_t j)^{-1} \sum_{\bar{U}_{tj}} 1_{\{k \in r\}} w_{Mkt} \|\mathbf{x}_k\| \|f(\cdot, \mathbf{x}_k)\|_A$$

und wegen $U_{tj} \subset U_t, \bar{U}_{tj} \subset U_t$ ist nach Voraussetzung

$$|\hat{X}_t(\lambda) - \hat{X}_t(\tilde{\lambda})| \leq h_j(\lambda, \tilde{\lambda}) N_t^{-1} \sum_{U_t} 1_{\{k \in r\}} w_{Mkt} + 2(N_t j)^{-1} \sum_{U_t} 1_{\{k \in r\}} w_{Mkt} \|\mathbf{x}_k\| f_s(\mathbf{x}_k) \quad (2.15)$$

Seien nun $\eta > 0$ und $\varepsilon > 0$ beliebig. Da $M_A = \sup N_t^{-1} \sum_{U_t} \|\mathbf{x}_k\| f_s(\mathbf{x}_k) < \infty$ ist, folgt aus der Markov-Ungleichung:

$$P\left(\left\{2(N_t j)^{-1} \sum_{U_t} 1_{\{k \in r\}} w_{Mkt} \|\mathbf{x}_k\| f_s(\mathbf{x}_k) > \eta/2\right\}\right) \leq \frac{4M_A}{j\eta}$$

Die rechte Seite der Ungleichung geht für $j \rightarrow \infty$ gegen Null. Also kann man $j \in \mathbb{N}$ so groß wählen, dass die rechte Seite der Ungleichung kleiner als $\varepsilon/2$ ist. Somit ist auch die linke Seite kleiner als $\varepsilon/2$. Für das gegebene $j \in \mathbb{N}$ gibt es ein $\delta > 0$, so dass $h_j(\lambda, \tilde{\lambda}) \leq \varepsilon\eta/4$ für alle $\lambda, \tilde{\lambda} \in A$ mit $\|\lambda - \tilde{\lambda}\| \leq \delta$ ist. Wiederum mit der Markov-Ungleichung ist

$$P\left(\left\{\sup_{\tilde{\lambda} \in A} \sup_{A(\tilde{\lambda}, \delta)} h_j(\lambda, \tilde{\lambda}) N_t^{-1} \sum_{U_t} 1_{\{k \in r\}} w_{Mkt} > \eta/2\right\}\right) \leq \varepsilon/2$$

Mit der obigen Ungleichung gibt es daher zu jedem $\eta > 0$ und jedem $\varepsilon > 0$ ein $\delta > 0$, so dass

$$P\left(\left\{\sup_{\tilde{\lambda} \in A} \sup_{A(\tilde{\lambda}, \delta)} |\hat{X}_t(\lambda) - \hat{X}_t(\tilde{\lambda})| > \eta\right\}\right) < \varepsilon$$

für alle $t \in \mathbb{N}$ gilt. □

Analog lässt sich zeigen, dass X_t gleichgradig stetig ist. Falls B kompakt ist, kann man auf die Schranke $\|f(\cdot, \mathbf{x}_k)\|_A \leq f_s(\mathbf{x}_k)$ verzichten. Denn für hinreichend großes $j \in \mathbb{N}$ enthält \bar{U}_{tj} für alle $t \in \mathbb{N}$ keine Einheiten.

Die obigen Ungleichung (2.15) lässt sich umformen:

$$\begin{aligned} |\hat{X}_t(\lambda) - \hat{X}_t(\tilde{\lambda})|^p &\leq 4^{p-1} \left\{ h_j(\lambda, \tilde{\lambda})^p N_t^{-p} \left| \sum_{U_t} 1_{\{k \in r\}} w_{Mkt} - N_t \right|^p \right. \\ &\quad + 2^p (N_t j)^{-p} \left| \sum_{U_t} 1_{\{k \in r\}} w_{Mkt} \|\mathbf{x}_k\| f_s(\mathbf{x}_k) - \sum_{U_t} \|\mathbf{x}_k\| f_s(\mathbf{x}_k) \right|^p \\ &\quad \left. + h_j(\lambda, \tilde{\lambda})^p + 2^p (N_t j)^{-p} \left(\sum_{U_t} \|\mathbf{x}_k\| f_s(\mathbf{x}_k) \right)^p \right\} \end{aligned}$$

Der letzte Term lässt sich über $j \in \mathbb{N}$ beliebig verkleinern. Für das gewählte $j \in \mathbb{N}$ findet sich dann eine Umgebung $A(\tilde{\lambda}, \delta)$, so dass auch $h_j(\lambda, \tilde{\lambda})^p$ eine vorgegeben Schranke nicht überschreitet. Die ersten beiden Terme auf der rechten Seite der Ungleichung konvergieren mit Annahme [A2] in Wahrscheinlichkeit gegen Null. Mit Annahme [A3] konvergieren sie für $p \geq 1$ auch in Erwartung gegen Null. Also ist \hat{X}_t im p -ten Mittel gleichgradige Stetigkeit für $p \geq 1$.

Zum Schluss dieses Abschnitts werden die Konsequenzen der Annahmen [A1], [A2] und [A3] auf $\hat{X}_t(\hat{\lambda}_t) - \hat{X}_t(\lambda_0)$ dargestellt. Hierzu ist $\sqrt{n_t}(\hat{\lambda}_t - \lambda_0) = O_p(1)$. Außerdem ist f stetig differenzierbar, wobei $f_1 : A \times B \rightarrow \mathbb{R}^d$ die Ableitung von f nach λ ist. Das Wegintegral für den Weg $\lambda_P(\lambda, a) = \lambda_0 + a(\lambda - \lambda_0)$ ist definiert als

$$F_1(\lambda_P(\lambda, \cdot), \cdot) = \int_0^1 f_1(\lambda_P(\lambda, a), \cdot) da$$

Die Anwendung des Mittelwertsatzes ergibt

$$\hat{X}_t(\hat{\lambda}_t) - \hat{X}_t(\lambda_0) = \mathbf{D}_t(\lambda_0)'(\hat{\lambda}_t - \lambda_0) + (\hat{\mathbf{D}}_t(\hat{\lambda}_t) - \mathbf{D}_t(\lambda_0))'(\hat{\lambda}_t - \lambda_0)$$

wobei $\mathbf{D}_t(\lambda) = N_t^{-1} \sum_U F_1(\lambda_P(\lambda, \cdot), \mathbf{x}_k)$ und

$$\hat{\mathbf{D}}_t(\lambda) = \frac{1}{N_t} \sum_{U_t} 1_{\{k \in r\}} w_{Mkt} F_1(\lambda_P(\lambda, \cdot), \mathbf{x}_k)$$

sind. Für $f_s : B \rightarrow \mathbb{R}^+$ mit $\|f_1(\cdot, \mathbf{x}_k)\|_A \leq f_s(\mathbf{x}_k)$ und Annahme [A1] ist $\hat{\mathbf{D}}_t$ mit dem obigen Satz

stochastisch gleichgradig stetig. Mit Annahme [A2] konvergiert $\hat{\mathbf{D}}_t$ punktweise in Wahrscheinlichkeit gegen \mathbf{D}_t . Mit der stochastischen Variante des Satzes von Arzelà-Ascoli konvergiert $\hat{\mathbf{D}}_t$ gegen \mathbf{D}_t auch stochastisch gleichmäßig auf jeder kompakten Menge. $\mathbf{D}_t(\lambda_0)$ ist beschränkt nach Annahme [A1]. Folglich konvergiert auch $\hat{X}_t(\hat{\lambda}_t) - \hat{X}_t(\lambda_0)$ in Wahrscheinlichkeit gegen Null und es gilt sogar $\sqrt{n_t}(\hat{X}_t(\hat{\lambda}_t) - \hat{X}_t(\lambda_0)) = O_p(1)$.

Falls der Parameter-Raum A kompakt ist, gilt mit der zusätzlichen Annahme [A3]:

$$\lim_{t \rightarrow \infty} \mathbb{E}[\|\hat{\mathbf{D}}_t(\hat{\lambda}_t) - \mathbf{D}_t(\lambda_0)\|^p] = 0$$

für $p \geq 1$. Damit konvergiert auch $\hat{\mathbf{D}}_t(\hat{\lambda}_t) - \mathbf{D}_t(\lambda_0)$ im p -ten Mittel gegen Null. Mit der Hölder-Ungleichung ist

$$\begin{aligned} \mathbb{E}[|\hat{X}_t(\hat{\lambda}_t) - \hat{X}_t(\lambda_0)|^p] &\leq 2^{p-1} \|\mathbf{D}_t(\lambda_0)\|^p \mathbb{E}[\|\hat{\lambda}_t - \lambda_0\|^p] + \\ &\quad 2^{p-1} \mathbb{E}[\|\hat{\mathbf{D}}_t(\hat{\lambda}_t) - \mathbf{D}_t(\lambda_0)\|^{2p}]^{1/2} \mathbb{E}[\|\hat{\lambda}_t - \lambda_0\|^{2p}]^{1/2} \end{aligned}$$

Da der Parameter-Raum kompakt ist, konvergiert $\mathbb{E}[\|\hat{\lambda}_t - \lambda_0\|^{2p}]$ gegen Null mit dem Satz von Lebesgue. Folglich konvergiert $\hat{X}_t(\hat{\lambda}_t) - \hat{X}_t(\lambda_0)$ im p -ten Mittel gegen Null, falls $\hat{\mathbf{D}}_t(\hat{\lambda}_t) - \mathbf{D}_t(\lambda_0)$ im $2p$ -ten Mittel konvergent ist.

3 Kalibration

Kalibration ist eine Methode um Schätzergebnisse unter Zuhilfenahme von Zusatzinformation zu verbessern. Die Zusatzinformation besteht darin, dass für bestimmte Merkmale in der Stichprobe die Gesamtwerte für die Grundgesamtheit bekannt sind. Mit Hilfe der Gesamtwerte lassen sich die Design-Gewichte so anpassen, dass der Schätzer auf Basis der Stichprobe mit den Werten in der Grundgesamtheit übereinstimmt. Hierbei sollen die angepassten Gewichte möglichst nahe an den ursprünglichen Design-Gewichte liegen, d.h. der Abstand zwischen den Design-Gewichten und den angepassten Gewichten soll minimal sein. Die angepassten Gewichte hängen also nicht nur von der zur Verfügung stehenden Zusatzinformation sondern auch von der gewählten Abstandsfunktion ab. Deville (1988) führte diese Idee als erstes mit einer quadratischen Abstandsfunktion ein. Der resultierende Schätzer ist der Generalised Regression Estimator (GREG), Cassel et al. (1976). Die Herleitung des Generalised Regression Estimator basierte ursprünglich auf einem linearen Zusammenhang zwischen den Hilfsmerkmalen und dem Untersuchungsmerkmal. Deville und Särndal (1992) zeigen, dass bei der Wahl anderer Abstandsfunktionen sich die daraus ergebenden Kalibrationsschätzer wie der Kalibrationsschätzer auf Basis der quadratischen Abstandsfunktion verhalten. Särndal und Swensson (1987) betrachten den Generalised Regression Estimator im Rahmen eines zweiphasigen Stichprobendesigns. Hierzu wird zwischen der Verfügbarkeit der vorhandenen Informationen auf der Basis der Grundgesamtheit und der ersten Phase unterschieden. Estevao und Särndal (2002) betrachten den Kalibrationsschätzer mit einer quadratischen Abstandsfunktion und untersuchen wie sich unterschiedliche Informationsmengen auf die Varianz des Kalibrationsschätzers auswirken.

Auf dieser Idee beruht der Kalibrationsansatz bei Nonresponse. Hierzu wird Nonresponse als eine zweite Auswahlphase betrachtet. Der Unterschied zum obigen Ansatz beruht darauf, dass die „Auswahlwahrscheinlichkeiten“ (Response-Wahrscheinlichkeiten) der zweiten Phase nicht bekannt sind. Aus diesem Grund wird der Kalibrationsansatz modifiziert, indem auf die Verwendung der Response-Wahrscheinlichkeiten verzichtet wird. Durch den Verzicht lassen sich keine Aussagen über die statistischen Eigenschaften des so gewonnenen Kalibrationsschätzers machen. Um die statistischen Eigenschaften bestimmen zu können, ist es notwendig ein Response-Modell anzunehmen. Hierbei ist zu beachten, dass die reziproken Response-Wahrscheinlichkeiten den im Kalibrationsansatz verwendeten Ausgleichsgewichten entsprechen müssen.

3.1 Kalibration bei Nonresponse

Ein einfacher Ansatz um Verzerrungen durch Nonresponse zu beheben wurde von Lundström und Särndal (1999) entwickelt. Bei diesem Ansatz wird nur auf Basis der Design-Gewichte kalibriert, d.h. die Response-Wahrscheinlichkeiten werden nicht berücksichtigt. Hieraus ergeben sich Ausgleichsgewichte, die eine mögliche Verzerrung durch Nonresponse kompensieren können. Theoretisch lässt sich die Verzerrung durch Nonresponse vollständig kompensieren, falls die unbekanntenen Response-Wahrscheinlichkeiten eine bestimmte funktionale Form aufweisen, siehe Proposition 4.3 in Lundström und Särndal (1999). Außerdem ist es notwendig, diese funktionale Form auch bei der Varianzschätzung anzunehmen, siehe Proposition 3.2 und 3.3 in Lundström und Särndal (1999). Allerdings basieren die Aussagen auf einem Vergleich mit dem Kalibrationsschätzer bei einem zweiphasigen Ziehungsverfahren und nicht auf dem asymptotischen Konzept von Fuller und Isaki (1981) und Isaki und Fuller (1982). Folgt man dem asymptotischen Konzept bei der Herleitung des Varianzschätzers, ergeben sich

Unterschiede in der Varianzformel. Das Ausmaß des Unterschieds hängt davon ab, ob das Untersuchungsmerkmal mit den Response-Wahrscheinlichkeiten nach Kontrolle bezüglich der Hilfsmerkmale korreliert ist, siehe unten. Im Folgenden wird der Kalibrationsansatz bei Nonresponse formal dargestellt.

Sei U eine Grundgesamtheit von Einheiten. Auf Basis eines Stichprobendesigns p_D werden Einheiten der Grundgesamtheit ausgewählt. Mit den ausgewählten Einheiten soll der Mittelwert Y für das Untersuchungsmerkmal y geschätzt werden. Der Horvitz-Thompson Schätzer \hat{Y}_{HT} lässt sich aufgrund von Nonresponse nicht berechnen. Jedoch gibt es Merkmale \mathbf{x} , deren Ausprägungen für alle Personen der Stichprobe $s \subset U$ zur Verfügung stehen. Für einige Merkmale, \mathbf{x}_A , kann sogar der Mittelwert in der Grundgesamtheit $\mathbf{X}_A = N^{-1} \sum_U \mathbf{x}_{Ak}$ vorhanden sein. Seien für jede Person $k \in U$ Merkmalsausprägungen \mathbf{x}_{Ak} und \mathbf{x}_{Bk} der Dimension d_A bzw. Dimension d_B gegeben. Der Vektor der Dimension $d = d_A + d_B$ ist

$$\mathbf{x}_k = \begin{pmatrix} \mathbf{x}_{Ak} \\ \mathbf{x}_{Bk} \end{pmatrix}$$

Falls keine Merkmal \mathbf{x}_A existiert, ist $\mathbf{x}_k = \mathbf{x}_{Bk}$. Weiterhin ist es möglich, dass für alle Merkmale der Mittelwert in der Grundgesamtheit existiert. In diesem Fall gilt $\mathbf{x}_k = \mathbf{x}_{Ak}$. Für das Fehlen von y_k für eine Einheit $k \in s$ wird von Quasi-Randomization wie in 2.2 ausgegangen. Hierbei ist das Fehlen von y unter den Einheiten unabhängig.

All diese Fälle lassen sich einheitlich darstellen. Sei $\hat{\mathbf{X}} : \Omega \rightarrow \mathbb{R}^d$ ein \mathcal{F} -messbarer Schätzer auf Basis der Merkmale \mathbf{x} . Die folgende Einteilung für die Möglichkeiten für $\hat{\mathbf{X}}$ erfolgt nach Estevao und Särndal (2002):

$$\text{Fall C1: } \hat{\mathbf{X}}_{HT} = \frac{1}{N} \sum_U 1_{\{k \in s\}} w_k \mathbf{x}_k$$

$$\text{Fall B2: } \mathbf{X} = \frac{1}{N} \sum_U \mathbf{x}_k$$

und

$$\text{Fall B1: } \hat{\mathbf{X}}_{B1} = \begin{pmatrix} \mathbf{X}_A \\ \hat{\mathbf{X}}_{BHT} \end{pmatrix}$$

$$\text{Fall A1: } \hat{\mathbf{X}}_{A1} = \begin{pmatrix} \mathbf{X}_A \\ \hat{\mathbf{X}}_{BCAL} \end{pmatrix}$$

wobei die einzelnen Größen von $\hat{\mathbf{X}}_{B1}$ und $\hat{\mathbf{X}}_{A1}$ durch

$$\hat{\mathbf{X}}_{BCAL} = \hat{\mathbf{X}}_{BHT} + \hat{\mathbf{B}}_A^{B'} (\mathbf{X}_A - \hat{\mathbf{X}}_{AHT})$$

$$\hat{\mathbf{B}}_A^B = (\sum_U 1_{\{k \in s\}} w_k \mathbf{x}_{Ak} \mathbf{x}'_{Ak})^{-1} \sum_U 1_{\{k \in s\}} w_k \mathbf{x}_{Ak} \mathbf{x}'_{Bk}$$

mit

$$\hat{\mathbf{X}}_{AHT} = N^{-1} \sum_U 1_{\{k \in s\}} w_k \mathbf{x}_{Ak}$$

$$\hat{\mathbf{X}}_{BHT} = N^{-1} \sum_U 1_{\{k \in s\}} w_k \mathbf{x}_{Bk}$$

gegeben sind. Die Fälle C1 und B2 sind Bestandteil der Arbeit von Lundström und Särndal (1999). Die Fälle A1 und B1 sind in Särndal und Lundström (2005) zu finden. Für den Fall A1 wird der Generalised Regression Estimator für die B Merkmale verwendet, siehe Särndal et al. (1992).

Beim Kalibrationsansatz für Nonresponse werden Kalibrationsgewichte $\hat{w}_{Mk}(s, r)$ auf Basis der Stichprobe s und der Einheiten in r so bestimmt, dass

$$\sum_U 1_{\{k \in r\}} (\hat{w}_{Mk} - w_k)^2 / 2w_k$$

unter der Nebenbedingung

$$\frac{1}{N} \sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_k = \hat{\mathbf{X}}$$

minimal wird. Die Lösung mit Hilfe des Lagrange-Ansatzes ergibt:

$$\hat{w}_{Mk} = w_k + N(\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{NR}})' (\sum_U 1_{\{k \in r\}} w_k \mathbf{x}_k \mathbf{x}_k')^{-1} w_k \mathbf{x}_k \quad (3.1)$$

wobei $\hat{\mathbf{X}}_{\text{NR}} = N^{-1} \sum_U 1_{\{k \in r\}} w_k \mathbf{x}_k$ ist. Der Kalibrationsansatz bei Nonresponse ist durch die Verwendung von (2.10) mit dem funktionalen Zusammenhang $f_M: A \times B \rightarrow [0, 1]$ mit

$$f_M(\lambda, \mathbf{x}_k) = 1 + \lambda' \mathbf{x}_k \quad (3.2)$$

und der Schätzfunktion $\hat{\lambda}: \Omega \rightarrow A$ mit

$$\hat{\lambda} = N(\sum_U 1_{\{k \in r\}} w_k \mathbf{x}_k \mathbf{x}_k')^{-1} (\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{NR}}) \quad (3.3)$$

definiert. Die Invertierbarkeit der Matrix $\sum_U 1_{\{k \in r\}} w_k \mathbf{x}_k \mathbf{x}_k'$ sei hierbei durch die folgende Annahme gegeben:

[R1] Für beliebige Transformationen $f: B \rightarrow \mathbb{R}^d$ sind die Eigenwerte von $N^{-1} \sum_U 1_{\{k \in r\}} f(\mathbf{x}_k) f(\mathbf{x}_k)'$ für jede Realisierung durch eine Konstante $M_E > 0$ gleichmäßig nach unten beschränkt und $\lim N^{-1} \sum_U f(\mathbf{x}_k) f(\mathbf{x}_k)'$ ist positiv definit.

Einsetzen von (3.2) und (3.3) in (2.10) ergibt den Kalibrationsschätzer

$$\hat{Y}_{\text{NRC}} = \hat{Y}_{\text{NR}} + (\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{NR}})' \hat{\mathbf{B}}_{\text{NR}} \quad (3.4)$$

wobei

$$\hat{\mathbf{B}}_{\text{NR}} = (\sum_U 1_{\{k \in r\}} w_k \mathbf{x}_k \mathbf{x}_k')^{-1} \sum_U 1_{\{k \in r\}} w_k \mathbf{x}_k y_k \text{ und}$$

$$\hat{Y}_{\text{NR}} = \frac{1}{N} \sum_U 1_{\{k \in r\}} w_k y_k$$

ist. Hierbei soll der Term $(\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{NR}})' \hat{\mathbf{B}}_{\text{NR}}$ für die mögliche Verzerrung kompensieren. Der nächste Abschnitt befasst sich mit der asymptotischen Varianz von (3.4).

3.2 Varianz des Kalibrationsschätzers

Der Kalibrationsschätzer \hat{Y}_{NRC} in der Form 3.4 lässt sich wie folgt umformen:

$$\hat{Y}_{\text{NRC}} = \hat{Y}_{\text{NR}} + (\hat{\mathbf{X}}_{\text{M}} - \hat{\mathbf{X}}_{\text{NR}})' \hat{\mathbf{B}}_{\text{NR}} + (\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{M}})' \hat{\mathbf{B}}_{\text{NR}} \quad (3.5)$$

wobei

$$\hat{\mathbf{X}}_{\text{M}} = \frac{1}{N} \sum_U 1_{\{k \in r\}} w_{\text{M}k} \mathbf{x}_k$$

mit $w_{\text{M}k} = w_k(1 + \lambda_0' \mathbf{x}_k)$ und λ_0 ein Element des Parameter-Raums $A \subset \mathbb{R}^d$ ist. Interessanterweise lässt sich $\lambda_0 \in A$ implizit berechnen:

$$\lambda_0 = N(\sum_U 1_{\{k \in r\}} w_k \mathbf{x}_k \mathbf{x}_k')^{-1} (\hat{\mathbf{X}}_{\text{M}} - \hat{\mathbf{X}}_{\text{NR}}) \quad (3.6)$$

Hierbei hängt λ_0 nicht vom gewählten $\hat{\mathbf{X}}$ ab! Einsetzen von λ_0 in die Gleichung (3.5) ergibt

$$\hat{Y}_{\text{NRC}} = \hat{Y}_{\text{NR}} + \frac{1}{N} \sum_U 1_{\{k \in r\}} w_k \lambda_0' \mathbf{x}_k y_k + (\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{M}})' \hat{\mathbf{B}}_{\text{NR}}$$

Die ersten beiden Terme sind in Summe:

$$\hat{Y}_{\text{M}} = \frac{1}{N} \sum_U 1_{\{k \in r\}} w_{\text{M}k} y_k = \frac{1}{N} \sum_U 1_{\{k \in r\}} w_k (1 + \lambda_0' \mathbf{x}_k) y_k$$

Dann ist

$$\hat{Y}_{\text{NRC}} = \hat{Y}_{\text{M}} + (\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{M}})' \hat{\mathbf{B}}_{\text{NR}}$$

oder auch

$$\hat{Y}_{\text{NRC}} = \hat{Y}_{\text{M}} + (\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{M}})' \mathbf{B}_{\text{NR}} + (\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{M}})' (\hat{\mathbf{B}}_{\text{NR}} - \mathbf{B}_{\text{NR}}) \quad (3.7)$$

wobei

$$\mathbf{B}_{\text{NR}} = (\sum_U (1 + \lambda_0' \mathbf{x}_k)^{-1} \mathbf{x}_k \mathbf{x}_k')^{-1} \sum_U (1 + \lambda_0' \mathbf{x}_k)^{-1} \mathbf{x}_k y_k$$

ist.

In den nachfolgenden Ausführungen wird die Voraussetzung der Proposition 4.3 bei Lundström und Särndal (1999) verwendet: es gibt ein $\lambda_0 \in A$, so dass $E[1_{\{k \in r\}} | \mathcal{F}] = \theta_k = (1 + \lambda_0' \mathbf{x}_k)^{-1}$ für alle $k \in U$ gilt. Damit wird das Response-Modell festgelegt. Und nur unter diesem Modell erhalten Lundström und Särndal (1999) einen nahezu erwartungstreuen Kalibrationsschätzer.²⁰ Damit gilt für

²⁰Särndal und Lundström (2005) betrachten zwar noch den Fall, dass ein perfekter linearer Zusammenhang zwischen der Untersuchungsvariablen und den Hilfsmerkmalen besteht, d.h. $y_k = \beta' \mathbf{x}_k$. Doch einerseits kann dieser Zusammenhang nicht gleichzeitig für alle möglichen Untersuchungsvariablen gelten. Und andererseits ist, da y für die Antwortausfälle unbeobachtbar ist, dieser Zusammenhang im Allgemeinen nicht überprüfbar. Zwar geben sie ein Verfahren an um diesen Zusammenhang zu überprüfen. Jedoch wird dabei das inhärenten Response-Modell benutzt.

die vier Fälle

$$\text{Fall C1: } E[\hat{\mathbf{X}}_{\text{HT}} - \hat{\mathbf{X}}_{\text{M}}|\mathcal{F}] = 0$$

$$\text{Fall B2: } E[\mathbf{X} - \hat{\mathbf{X}}_{\text{M}}|\mathcal{F}] = \mathbf{X} - \hat{\mathbf{X}}_{\text{HT}}$$

und

$$\text{Fall B1: } E[\hat{\mathbf{X}}_{\text{B1}} - \hat{\mathbf{X}}_{\text{M}}|\mathcal{F}] = \begin{pmatrix} \mathbf{X}_{\text{A}} - \hat{\mathbf{X}}_{\text{AHT}} \\ 0 \end{pmatrix}$$

$$\text{Fall A1: } E[\hat{\mathbf{X}}_{\text{A1}} - \hat{\mathbf{X}}_{\text{M}}|\mathcal{F}] = \hat{\mathbf{B}}'_{\text{A}}(\mathbf{X}_{\text{A}} - \hat{\mathbf{X}}_{\text{AHT}}) = \mathbf{B}'_{\text{A}}(\mathbf{X}_{\text{A}} - \hat{\mathbf{X}}_{\text{AHT}}) + (\hat{\mathbf{B}}_{\text{A}} - \mathbf{B}_{\text{A}})'(\mathbf{X}_{\text{A}} - \hat{\mathbf{X}}_{\text{AHT}})$$

mit $\hat{\mathbf{B}}_{\text{A}} = (\mathbf{I}|\hat{\mathbf{B}}_{\text{A}}^{\text{B}})$ und $\mathbf{B}_{\text{A}} = (\sum_U \mathbf{x}_{\text{Ak}}\mathbf{x}'_{\text{Ak}})^{-1} \sum_U \mathbf{x}_{\text{Ak}}\mathbf{x}'_{\text{Ak}}$. \mathbf{I} ist die d_{A} -dimensionale Einheitsmatrix. Außerdem gilt für alle Fälle, dass mit den Annahmen [A2] und [R1] $\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{M}} = O_p(n^{-1/2})$ gilt. Mit der zusätzlichen Annahme [A3] ist das $2p$ -te Mittel $O(n^{-p})$.

Wenn jetzt noch $\hat{\mathbf{B}}_{\text{NR}} - \mathbf{B}_{\text{NR}} = O_p(n^{-1/2})$ gelten würde, dann wäre der letzten Term in (3.7) $O_p(n^{-1})$. Unter Vernachlässigung des letzten Terms lässt sich dann die Varianz approximieren. Sei

$$\hat{\mathbf{T}}_{\text{M}} = \sum_U 1_{\{k \in r\}} w_{\text{Mk}} \mathbf{x}_k \mathbf{x}'_k \text{ und}$$

$$\hat{\mathbf{T}}_{\text{NR}} = \sum_U 1_{\{k \in r\}} w_{\text{Mk}} \mathbf{x}_k \mathbf{x}'_k = \sum_U 1_{\{k \in r\}} w_{\text{Mk}} (1 + \lambda'_0 \mathbf{x}_k)^{-1} \mathbf{x}_k \mathbf{x}'_k$$

Mit der Jensenschen Ungleichung, gilt für jedes $\nu \in \mathbb{R}^d$ mit $\|\nu\| = 1$

$$N^{-1} \nu' \hat{\mathbf{T}}_{\text{NR}} \nu \geq \frac{(N^{-1} \nu' \hat{\mathbf{T}}_{\text{M}} \nu)^2}{N^{-1} \sum_U 1_{\{k \in r\}} w_{\text{Mk}} (\nu' \mathbf{x}_k)^2 (1 + \lambda'_0 \mathbf{x}_k)}$$

siehe Anhang A.1. Wählt man für ν den Eigenvektor zum kleinsten Eigenwert, so ist

$$\|N^{-1} \hat{\mathbf{T}}_{\text{NR}}^{-1}\| = \lambda_{\min}(N^{-1} \hat{\mathbf{T}}_{\text{NR}})^{-1} \leq \|N^{-1} \sum_U 1_{\{k \in r\}} (1 + \lambda'_0 \mathbf{x}_k) w_{\text{Mk}} \mathbf{x}_k \mathbf{x}'_k\| \|N \hat{\mathbf{T}}_{\text{M}}^{-1}\|^2$$

da $\lambda_{\max}(N^{-1} \hat{\mathbf{T}}_{\text{NR}}^{-1}) = \lambda_{\min}(N^{-1} \hat{\mathbf{T}}_{\text{NR}})^{-1}$ ist. Mit den Annahmen [A1], [A2] und [R1] ist die rechte Seite der Ungleichung $O_p(1)$ und damit auch die linke Seite der Ungleichung. Mit den Annahmen [A1] und [R1] ist der Populationswert $(N^{-1} \sum_U (1 + \lambda'_0 \mathbf{x}_k)^{-1} \mathbf{x}_k \mathbf{x}'_k)^{-1} = O(1)$. Somit ist auch $\mathbf{B}_{\text{NR}} = O(1)$. Mit $e_k^* = y_k - \mathbf{x}'_k \mathbf{B}_{\text{NR}}$ ist

$$\hat{\mathbf{B}}_{\text{NR}} - \mathbf{B}_{\text{NR}} = (\sum_U 1_{\{k \in r\}} w_{\text{Mk}} \mathbf{x}_k \mathbf{x}'_k)^{-1} \sum_U 1_{\{k \in r\}} (1 + \lambda'_0 \mathbf{x}_k)^{-1} w_{\text{Mk}} \mathbf{x}_k e_k^*$$

und somit gilt $\hat{\mathbf{B}}_{\text{NR}} - \mathbf{B}_{\text{NR}} = O_p(n^{-1/2})$. Mit der zusätzlichen Annahme [A3] und Annahme [R1] ist das $2p$ -te Mittel $O(n^{-p})$. Mit 3.7 ist daher

$$\hat{Y}_{\text{NRC}} = \hat{Y}_{\text{M}} + (\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{M}})' \mathbf{B}_{\text{NR}} + O_p(n^{-1}) \quad (3.8)$$

Somit gilt der folgende Satz.

Satz (Kalibration bei Nonresponse). Mit den Response-Wahrscheinlichkeiten $\theta_k = (1 + \lambda'_0 \mathbf{x}_k)^{-1}$ ist bei Quasi-Randomization und den Annahmen [A1], [A2], [A3] und [R1] die Verzerrung von \hat{Y}_{NRC}

$$E[\hat{Y}_{\text{NRC}} - Y] = O(n^{-1}) \quad (3.9)$$

und die Varianz von \hat{Y}_{NRC} ist

$$V[\hat{Y}_{\text{NRC}}] = \frac{1}{N^2} \sum \sum_U \Delta_{kl} \frac{e_{1k}^*}{\pi_k} \frac{e_{1l}^*}{\pi_l} + \frac{1}{N^2} E_p \left[\sum \sum_U 1_{\{k,l \in s\}} \Delta_{Mkl} (w_{Mk} e_k^*) (w_{Ml} e_l^*) \right] + O(n^{-2}) \quad (3.10)$$

mit $\Delta_{Mkk} = \theta_k(1 - \theta_k)$, $\Delta_{Mkl} = 0$ und

$$\text{Fall C1: } e_{1k}^* = y_k$$

$$\text{Fall B2: } e_{1k}^* = y_k - \mathbf{x}'_k \mathbf{B}_{\text{NR}}$$

$$\text{Fall B1: } e_{1k}^* = y_k - \mathbf{x}'_{Ak} \mathbf{B}_{\text{NR}}^A$$

$$\text{Fall A1: } e_{1k}^* = y_k - \mathbf{x}'_{Ak} \mathbf{B}_A \mathbf{B}_{\text{NR}}$$

\mathbf{B}_{NR}^A sind die ersten d_A Einträge des Vektors \mathbf{B}_{NR} . □

Die Varianz wird geschätzt, indem e_k^* durch $\hat{e}_k^* = y_k - \mathbf{x}'_k \hat{\mathbf{B}}_{\text{NR}}$, w_{Mk} durch \hat{w}_{Mk} und Δ_{Mkk} durch $\hat{\Delta}_{Mkk} = \hat{\theta}_k(1 - \hat{\theta}_k)$ ersetzt wird. Der Schätzer 3.3 wird für die geschätzten Response-Wahrscheinlichkeiten verwendet. Ein Varianzschätzer ist damit

$$\begin{aligned} \hat{V}_{\text{NRC}} &= \frac{1}{N^2} \sum \sum_U 1_{\{k,l \in r\}} \frac{\Delta_{kl}}{\pi_{kl}} (\hat{w}_{Mk} \hat{e}_{1k}^*) (\hat{w}_{Ml} \hat{e}_{1l}^*) - \frac{1}{N^2} \sum_U 1_{\{k \in r\}} (1 - \pi_k) (1 - \hat{\theta}_k) (\hat{w}_{Mk} \hat{e}_{1k}^*)^2 \\ &\quad + \frac{1}{N^2} \sum_U 1_{\{k \in r\}} (1 - \hat{\theta}_k) (\hat{w}_{Mk} \hat{e}_k^*)^2 \end{aligned} \quad (3.11)$$

wobei

$$\text{Fall C1: } \hat{e}_{1k}^* = y_k$$

$$\text{Fall B2: } \hat{e}_{1k}^* = y_k - \mathbf{x}'_k \hat{\mathbf{B}}_{\text{NR}}$$

$$\text{Fall B1: } \hat{e}_{1k}^* = y_k - \mathbf{x}'_{Ak} \hat{\mathbf{B}}_{\text{NR}}^A$$

$$\text{Fall A1: } \hat{e}_{1k}^* = y_k - \mathbf{x}'_{Ak} \hat{\mathbf{B}}_A \hat{\mathbf{B}}_{\text{NR}}$$

benutzt wird. Es ist $\hat{\mathbf{B}}_A = (\sum_U 1_{\{k \in s\}} w_k \mathbf{x}_{Ak} \mathbf{x}'_{Ak})^{-1} \sum_U 1_{\{k \in s\}} w_k \mathbf{x}_{Ak} \mathbf{x}'_k$ und $\hat{\mathbf{B}}_{\text{NR}}^A$ sind die ersten d_A Einträge von $\hat{\mathbf{B}}_{\text{NR}}$

Sei $\mathbf{B} = (\sum_U \mathbf{x}_k \mathbf{x}'_k)^{-1} \sum_U \mathbf{x}_k y_k$ der Kleinste-Quadrate-Schätzer, dann lässt sich (3.8) umformen zu:

$$\hat{Y}_{\text{NRC}} = \hat{Y}_M + (\hat{\mathbf{X}} - \hat{\mathbf{X}}_M)' \mathbf{B} + (\hat{\mathbf{X}} - \hat{\mathbf{X}}_M)' (\mathbf{B}_{\text{NR}} - \mathbf{B}) + O_p(n^{-1})$$

Es gilt

$$\mathbf{B}_{\text{NR}} - \mathbf{B} = (\sum_U \theta_k \mathbf{x}_k \mathbf{x}'_k)^{-1} \sum_U \theta_k \mathbf{x}_k (y_k - \mathbf{x}'_k \mathbf{B})$$

Analog zu Bethlehem (1988) ist $\mathbf{B}_{\text{NR}} = \mathbf{B}$, falls die Korrelation zwischen den mit \mathbf{x}_k multiplizierten Residuen $y_k - \mathbf{x}'_k \mathbf{B}$ aus der Kleinste-Quadrate-Schätzung auf Populationsebene unkorreliert zum Response-Verhalten ist:

$$\frac{1}{N} \sum_U \theta_k \mathbf{x}_k (y_k - \mathbf{x}'_k \mathbf{B}) = \frac{1}{N} \sum_U (\theta_k - \theta_{\text{NR}}) \mathbf{x}_k (y_k - \mathbf{x}'_k \mathbf{B})$$

wobei $\theta_{\text{NR}} = N^{-1} \sum_U \theta_k$ die mittlere Response-Wahrscheinlichkeit ist. Die rechte Seite dieser Gleichung ist die Kovarianz zwischen den Response-Wahrscheinlichkeiten und den Werten $\mathbf{x}_k(y_k - \mathbf{x}'_k \mathbf{B})$. Wenn aber die Hilfsmerkmale die Untersuchungsvariable gut erklären können, wird der Term $y_k - \mathbf{x}'_k \mathbf{B}$ klein. Der Gesamtausdruck hängt aber noch von zwei weiteren Termen, \mathbf{x}_k und $\theta_k - \theta_{\text{NR}}$, ab. Allerdings heben sich diese beiden Ausdrücke auf, da große \mathbf{x}_k kleine Response-Wahrscheinlichkeiten unter dem gegebenen Modell implizieren.

Bei Särndal und Lundström (2005) wird betont, dass sich \mathbf{B}_{NR} und \mathbf{B} beachtlich unterscheiden können. Dennoch verwenden Lundström und Särndal (1999) das Stichprobenpendant von \mathbf{B} bei ihrer Varianzformel.

3.3 Im Vergleich zu Lundström und Särndal (1999)

Die Varianzformel für \hat{Y}_{NRC} im Fall von $\hat{\mathbf{X}} = \hat{\mathbf{X}}_{\text{HT}}$ basiert bei Lundström und Särndal (1999) auf der Annahme, dass der Bias $E[\hat{Y}_{\text{NRC}} - Y | \mathcal{F}] \approx 0$ ist. Deshalb sei kurz die Analyse des Bias vorangestellt.

Der Kalibrationsschätzer (3.4) lässt sich auch direkt für $\hat{\mathbf{X}} = \hat{\mathbf{X}}_{\text{HT}}$ zu

$$\hat{Y}_{\text{NRS}} = \hat{Y}_{\text{NR}} + (\hat{\mathbf{X}}_{\text{HT}} - \hat{\mathbf{X}}_{\text{NR}})' \mathbf{B}_{\text{NR}} + (\hat{\mathbf{X}}_{\text{HT}} - \hat{\mathbf{X}}_{\text{NR}})' (\hat{\mathbf{B}}_{\text{NR}} - \mathbf{B}_{\text{NR}})$$

umformen. Lundström und Särndal (1999) vernachlässigen den letzten Term mit der Begründung, dass $E[\hat{\mathbf{B}}_{\text{NR}} - \mathbf{B}_{\text{NR}}] \approx 0$ ist. Allerdings ist der Term $\hat{\mathbf{X}}_{\text{HT}} - \hat{\mathbf{X}}_{\text{NR}}$, der den Bias ausgleichen soll, von Null verschieden. Mit der Hölder-Ungleichung ist der letzte Term in Erwartung Null, falls das quadratische Mittel von $\hat{\mathbf{B}}_{\text{NR}} - \mathbf{B}_{\text{NR}}$ Null ist.²¹ Damit ist der Bias in ihrer Proposition 4.1

$$E[\hat{Y}_{\text{NRS}} - Y] \approx -\frac{1}{N} \sum_U (1 - \theta_k)(y_k - \mathbf{x}'_k \mathbf{B}_{\text{NR}})$$

Man beachte hierbei, dass \mathbf{B}_{NR} nicht der Kleinste-Quadrate-Schätzer einer Regression von y auf \mathbf{x} in der Population U ist. Außerdem ist der bedingte Bias durch

$$E[\hat{Y}_{\text{NRS}} - \hat{Y}_{\text{HT}} | \mathcal{F}] \approx -\frac{1}{N} \sum_U 1_{\{k \in s\}} w_k (1 - \theta_k)(y_k - \mathbf{x}'_k \mathbf{B}_{\text{NR}})$$

gegeben. Unter der Annahme, dass die Response-Wahrscheinlichkeiten durch $\theta_k = (1 + \lambda'_0 \mathbf{x}_k)^{-1}$ für $k \in U$ und fixes $\lambda_0 \in A$ gegeben sind, ist nach Definition $\theta_k(1 + \lambda'_0 \mathbf{x}_k) = 1$ und somit gilt $1 - \theta_k = \theta_k \lambda'_0 \mathbf{x}_k$. Damit ist der approximative Bias, Proposition 4.3 in Lundström und Särndal (1999),

$$E[\hat{Y}_{\text{NRS}} - Y] \approx -\sum_U (1 - \theta_k)(y_k - \mathbf{x}'_k \mathbf{B}_{\text{NR}}) = -\lambda'_0 \sum_U \theta_k \mathbf{x}_k y_k + \lambda'_0 \sum_U \theta_k \mathbf{x}_k \mathbf{x}'_k \mathbf{B}_{\text{NR}} = 0$$

Dieser Bias von Null ergibt sich nur unter der Annahme, dass das Response-Modell die obige funktionale Form hat.

Bei der Varianz argumentieren Lundström und Särndal (1999) wie folgt. Ohne Nonresponse, d.h. $r = s$, ist sowohl $\hat{Y}_{\text{NR}} = \hat{Y}_{\text{HT}}$ und $\hat{\mathbf{X}}_{\text{NR}} = \hat{\mathbf{X}}_{\text{HT}}$. Somit ist ohne Nonresponse und unter Verwendung von (3.4), $\hat{Y}_{\text{NRS}} = \hat{Y}_{\text{HT}}$. Wählt man als Vergleichskriterium den Horvitz-Thompson Schätzer, dann ist der mittlere quadratische Fehler von \hat{Y}_{NRS}

$$\text{MSE}[\hat{Y}_{\text{NRS}}] = V[\hat{Y}_{\text{HT}}] + E[V[\hat{Y}_{\text{NRS}} | \mathcal{F}]] + 2\text{Cov}[\hat{Y}_{\text{HT}}, E[\hat{Y}_{\text{NRS}} - \hat{Y}_{\text{HT}} | \mathcal{F}]] + E[E[\hat{Y}_{\text{NRS}} - \hat{Y}_{\text{HT}} | \mathcal{F}]^2]$$

²¹Im letzten Abschnitt wurde gezeigt, dass $\hat{\mathbf{B}}_{\text{NR}} - \mathbf{B}_{\text{NR}} = O_p(n^{-1/2})$ ist.

Falls der bedingte Bias Null ist, dann ist der mittlere quadratische Fehler

$$\text{MSE}[\hat{Y}_{\text{NRS}}] = \text{V}[\hat{Y}_{\text{HT}}] + \text{E}[\text{V}[\hat{Y}_{\text{NRS}}|\mathcal{F}]]$$

Unter Verwendung von

$$\hat{\lambda} = N(\sum_U 1_{\{k \in r\}} w_k \mathbf{x}_k \mathbf{x}_k')^{-1} (\hat{\mathbf{X}}_{\text{HT}} - \hat{\mathbf{X}}_{\text{NR}})$$

in (2.10) ist $\hat{\mathbf{X}}_{\hat{M}} = \hat{\mathbf{X}}_{\text{HT}}$ und $\hat{Y}_{\hat{M}} = \hat{Y}_{\text{NRS}} + (\hat{\mathbf{X}}_{\text{HT}} - \hat{\mathbf{X}}_{\text{NR}})' \hat{\mathbf{B}}_{\hat{M}}$ und daher gilt

$$\hat{Y}_{\text{NRS}} = \hat{Y}_{\hat{M}} + (\hat{\mathbf{X}}_{\text{HT}} - \hat{\mathbf{X}}_{\hat{M}})' \hat{\mathbf{B}}_{\hat{M}}$$

Dies ist der Kalibrationsschätzer, falls man die Gewichte $\hat{w}_{Mk} = (\pi_k \hat{\theta}_k)^{-1}$, $\hat{\theta}_k = (1 + \hat{\lambda}' \mathbf{x}_k)^{-1}$, mit Hilfe von $\hat{\mathbf{X}}_{\text{HT}}$ kalibriert. Sei

$$\hat{Y}_{\text{CAL}} = \hat{Y}_{\hat{M}} + (\hat{\mathbf{X}}_{\text{HT}} - \hat{\mathbf{X}}_{\hat{M}})' \mathbf{B} + (\hat{\mathbf{X}}_{\text{HT}} - \hat{\mathbf{X}}_{\hat{M}})' (\hat{\mathbf{B}}_{\hat{M}} - \mathbf{B}) \quad (3.12)$$

mit

$$\begin{aligned} \mathbf{B} &= (\sum_U \mathbf{x}_k \mathbf{x}_k')^{-1} \sum_U \mathbf{x}_k y_k \\ \hat{\mathbf{B}}_{\hat{M}} &= (\sum_U 1_{\{k \in r\}} w_{Mk} \mathbf{x}_k \mathbf{x}_k')^{-1} \sum_U 1_{\{k \in r\}} w_{Mk} \mathbf{x}_k y_k \end{aligned}$$

der Kalibrationsschätzer, falls man die Gewichte $w_{Mk} = (\pi_k \theta_k)^{-1}$ unter Zuhilfenahme von $\hat{\mathbf{X}}_{\text{HT}}$ kalibriert. Dieser hat unter den Annahmen [A1]-[A3] und [R1] die Varianz

$$\text{V}[\hat{Y}_{\text{CAL}}] = \frac{1}{N^2} \sum \sum_U \Delta_{kl} \frac{y_k}{\pi_k} \frac{y_l}{\pi_l} + \frac{1}{N^2} \text{E}_p \left[\sum \sum_U 1_{\{k, l \in s\}} \Delta_{Mkl} (w_{Mk} e_k) (w_{Ml} e_l) \right] + O(n^{-2}) \quad (3.13)$$

mit $e_k = y_k - \mathbf{x}_k' \mathbf{B}$. Vergleicht man den mittleren quadratischen Fehler mit der asymptotischen Varianz von \hat{Y}_{CAL} , dann haben, falls der Bias Null ist, \hat{Y}_{NRS} und \hat{Y}_{CAL} beide dieselbe erste Varianzkomponente. Mit dieser Rechtfertigung verwenden Lundström und Särndal (1999) als Proxy für $\text{V}[\hat{Y}_{\text{NRS}}]$ die asymptotische Varianz (3.13). Vergleicht man (3.7) mit (3.12), dann wird implizit behauptet, dass $\hat{\mathbf{B}}_{\text{NR}}$ durch $\hat{\mathbf{B}}_{\hat{M}}$ ausgetauscht werden kann.

Zusammenfassend lässt sich sagen, dass es zwei kritische Punkte bei der Kalibration gibt. Zum einen können die Ausgleichsgewichte negativ werden und zum anderen berücksichtigt die Varianzformel bei Lundström und Särndal (1999) nicht den Anteil, der durch die Schätzung der Ausgleichsgewichte erzeugt wird.

4 Propensity Score Adjustment

Eine Alternative zur Kalibration besteht in der Verwendung des Propensity Score adjustierten (PSA) Schätzers. Der PSA-Schätzer basiert auf den geschätzten Response-Wahrscheinlichkeiten aus einem statistischen Modell (z.B. Logistisches Regressionsmodell). Hierzu werden die reziproken geschätzten Response-Wahrscheinlichkeiten als Ausgleichsgewichte verwendet. Dabei ist zu beachten, dass durch die Schätzung der Response-Wahrscheinlichkeiten zusätzliche Variation generiert wird, die bei der Schätzung der Varianz berücksichtigt werden muss.

Im Folgenden wird die Herleitung des PSA-Schätzers dargestellt. Sei U eine Grundgesamtheit von Einheiten. Auf Basis eines Stichprobendesigns p_D werden Einheiten $k \in U$ mit dem Ziel das Untersuchungsmerkmal y nebst weiteren Hilfsvariablen \mathbf{x} zu erheben ausgewählt. Aufgrund von Non-response wird für einige Einheiten der Stichprobe das Untersuchungsmerkmal y nicht beobachtet. Sei $1_{\{k \in r\}} : \Omega \rightarrow \{0, 1\}$, die Zufallsgröße, die angibt, ob das Untersuchungsmerkmal y für eine Einheit beobachtet wird. Der Mittelwert des Untersuchungsmerkmal y in der Grundgesamtheit wird geschätzt, indem die Response-Wahrscheinlichkeiten $\theta_k : \Omega \rightarrow [0, 1]$ für alle Einheiten der Stichprobe modelliert und geschätzt werden.

Für die Modellierung und Schätzung der Response-Wahrscheinlichkeiten wird üblicherweise ein logistisches Regressionsmodell verwendet. Hierdurch wird sichergestellt, dass die geschätzten Response-Wahrscheinlichkeiten und daraus abgeleiteten Ausgleichsgewichte positiv sind. Dies ist ein Vorteil gegenüber dem Kalibrationsansatz, bei dem es zu negativen Gewichten kommen kann. Dieses Problem wird bei der Kalibration meistens ignoriert. Ein alternative Ansatz besteht darin, die Hilfsmerkmale, die zu negativen Gewichten führen, aus der Kalibration zu entfernen.

Das logistische Regressionsmodell gehört zu der Klasse der Verallgemeinerten Linearen Modelle, Nelder und Wedderburn (1972). Ein Verallgemeinertes Lineares Modell besteht aus zwei Komponenten:

- (1) Die Zufallskomponente spezifiziert die Verteilung der Zielvariablen, zum Beispiel binomial.
- (2) Die systematische Komponente spezifiziert, wie der (bedingte) Erwartungswert der Zielvariablen von den Hilfsmerkmalen abhängt.

Hierbei beeinflussen die Merkmale \mathbf{x} die Verteilung der Zufallskomponenten durch eine lineare Funktion (Linearer Prädiktor). Die Link-Funktion verknüpft den Erwartungswert oder eine Transformation des Erwartungswertes mit dem linearen Prädiktor. Im Falle der Bernoulli-Verteilung ist die Wahrscheinlichkeitsfunktion durch

$$q(a_k, \psi_k) = \exp \{ a_k \psi_k - b(\psi_k) \}, \quad a_k \in \{0, 1\}$$

gegeben, wobei $b : \mathbb{R} \rightarrow \mathbb{R}^+$ eine Funktion definiert als $b(\cdot) = \log\{1 + \exp(\cdot)\}$ ist und der transformierte Erwartungswert

$$\psi_k = \log \left\{ \frac{\theta_k}{1 - \theta_k} \right\} = f_L(\lambda_0' \mathbf{x}_k) \tag{4.1}$$

durch die dreimal stetig differenzierbare und injektive Funktion $f_L : C \rightarrow \mathbb{R}$, $C \subset \mathbb{R}$ mit dem linearen Term $\lambda_0' \mathbf{x}_k$, $\lambda_0 \in A \subset \mathbb{R}^d$, verknüpft ist. Die Funktion b ist beliebig oft stetig differenzierbar. Sei b_i die i -te Ableitung, dann gilt $b_1(\psi_k) = \theta_k$ und $b_2(\psi_k) = \theta_k(1 - \theta_k)$.

Der Parameter $\lambda_0 \in A$ lässt sich mit Hilfe der Maximum-Likelihood Methode schätzen, indem die Log-Likelihood-Funktion

$$L(\lambda) = \sum_U 1_{\{k \in r\}} f_L(\lambda' \mathbf{x}_k) - \sum_U 1_{\{k \in s\}} b(f_L(\lambda' \mathbf{x}_k)) \quad (4.2)$$

bezüglich λ maximiert wird. Sei $f_M(\lambda, \mathbf{x}_k) = f_N(\lambda, \mathbf{x}_k)^{-1}$ wobei $f_N(\lambda, \mathbf{x}_k) = b_1(f_L(\lambda' \mathbf{x}_k))$ ist. Mit dem gewählten Modell und dem geschätzten Parameter $\hat{\lambda}$ lassen sich die Gewichte w_{Mk} durch $\hat{w}_{Mk} = w_k f_M(\hat{\lambda}, \mathbf{x}_k)$ schätzen und (2.10) anwenden.

Allerdings ist durch die Verwendung des Modells im Design-basierten Ansatz nicht sichergestellt, ob es für jede mögliche Stichprobe eine Maximum-Likelihood-Lösung existiert. Deshalb wird in Abschnitt 4.2 gezeigt, dass es mit einer Restriktion an das Stichprobendesign asymptotisch eine Lösung mit Wahrscheinlichkeit Eins gibt. Dabei spielt die Fisher-Information eine zentrale Rolle. Durch die Restriktion ist die Inverse der Fisher-Information $O_p(n^{-1})$, während eine gewichtete Version der Inversen unter den in Abschnitt 1.3 gesetzten Annahmen $O_p(1)$ ist. Deshalb werden Eigenschaften der Fisher-Information in Abschnitt 4.1 dargestellt.

Die Haupt-Schwierigkeit beim PSA-Schätzer liegt darin, dass $\hat{\lambda} - \lambda_0$ durch eine asymptotisch gleichwertige Größe, für die sich Erwartungswert und Varianz berechnen lässt, ersetzt werden kann. Die asymptotische gleichwertige Größe ist das Produkt aus der inversen Fisher-Information und der Score-Funktion, welches die erste Ableitung der Log-Likelihood-Funktion ist. Dies ist Gegenstand von Abschnitt 4.3. Darauf basierend lässt sich ein einfacher Varianzschätzer für den PSA-Schätzer, ohne Kalibration (Abschnitt 4.4) und mit Kalibration (Abschnitt 4.5), herleiten.

Die asymptotischen Eigenschaften des PSA-Schätzers wurde bereits von Kim und Kim (2007) und Kim und Riddles (2012) untersucht. Dabei wird ohne Beweis die Approximation von $\hat{\lambda} - \lambda_0$ durch das Produkt aus inverser Fisher-Information und Score-Funktion benutzt. Außerdem hat die asymptotische Varianz in Abschnitt 4.4 eine einfachere Darstellung als bei den genannten Autoren und sie ist mit Annahme [A3] bis auf Terme der Ordnung $O(n^{-2})$ genau. Außerdem arbeiten die Autoren unter der Annahme, dass Funktionen in Abhängigkeit von λ gleichmäßig konvergieren. Diese Annahme ist nicht notwendig, siehe Abschnitt 2.

4.1 Score-Funktion und Fisher-Information

Die Score-Funktion ist definiert als die erste Ableitung der Likelihood-Funktion. Mit $\mathbf{h}_1(\lambda, \mathbf{x}_k) = \mathbf{x}_k f_{L1}(\lambda' \mathbf{x}_k)$, ist die erste Ableitung von (4.2)

$$\mathbf{u}(\lambda) = \sum_U 1_{\{k \in r\}} \mathbf{h}_1(\lambda, \mathbf{x}_k) - \sum_U 1_{\{k \in s\}} \mathbf{h}_1(\lambda, \mathbf{x}_k) f_N(\lambda, \mathbf{x}_k) \quad (4.3)$$

wobei f_{L1} die erste Ableitung von f_L ist. Für $f_L = \text{id}$ (Identität) vereinfacht sich die Score-Funktion zu

$$\mathbf{u}(\lambda) = \sum_U 1_{\{k \in r\}} \mathbf{x}_k - \sum_U 1_{\{k \in s\}} \mathbf{x}_k f_N(\lambda, \mathbf{x}_k)$$

Sei $\lambda_0 \in A$ der zu schätzende Parameter, dann ist die Score-Funktion ausgewertet an der Stelle λ_0 gegeben durch $\mathbf{u}_M = \mathbf{u}(\lambda_0)$. Für \mathbf{u}_M gilt

$$E[\mathbf{u}_M | \mathcal{F}] = 0$$

und die bedingte Varianz ist gegeben durch

$$\hat{\mathbf{F}}_{\mathbf{M}} = \mathbf{V}[\mathbf{u}_{\mathbf{M}}|\mathcal{F}] = \sum_U 1_{\{k \in s\}} f_{\mathbf{N}_1}(\lambda_0, \mathbf{x}_k) \mathbf{h}_1(\lambda_0, \mathbf{x}_k) \mathbf{h}_1(\lambda_0, \mathbf{x}_k)'$$

wobei $f_{\mathbf{N}_1}(\lambda, \mathbf{x}_k) = b_2(f_L(\lambda' \mathbf{x}_k))$ ist. Die bedingte Varianz als Funktion von λ sei:

$$\hat{\mathbf{F}}(\lambda) = \sum_U 1_{\{k \in s\}} f_{\mathbf{N}_1}(\lambda, \mathbf{x}_k) \mathbf{h}_1(\lambda, \mathbf{x}_k) \mathbf{h}_1(\lambda, \mathbf{x}_k)' \quad (4.4)$$

$\hat{\mathbf{F}}_{\mathbf{M}} = \hat{\mathbf{F}}(\lambda_0)$ sei die Fisher-Information ausgewertet an der Stelle $\lambda_0 \in A$. Für die negativen zweiten Ableitungen der Likelihood-Funktion gilt

$$\begin{aligned} \hat{\mathbf{H}}(\lambda) &= \hat{\mathbf{F}}(\lambda) + \hat{\mathbf{Q}}(\lambda) \text{ mit} \\ \hat{\mathbf{Q}}(\lambda) &= \sum_U 1_{\{k \in s\}} f_{\mathbf{N}}(\lambda, \mathbf{x}_k) \mathbf{h}_2(\lambda, \mathbf{x}_k) - \sum_U 1_{\{k \in r\}} \mathbf{h}_2(\lambda, \mathbf{x}_k) \end{aligned} \quad (4.5)$$

und $\mathbf{h}_2(\lambda, \mathbf{x}_k) = \mathbf{x}_k \mathbf{x}_k' f_{L2}(\lambda' \mathbf{x}_k)$, wobei f_{L2} die zweite Ableitung von f_L ist. Mit $f_L = \text{id}$ ist $\hat{\mathbf{H}} = \hat{\mathbf{F}}$. Die negativen zweiten Ableitungen ausgewertet an der Stelle $\lambda_0 \in A$ seien $\hat{\mathbf{H}}_{\mathbf{M}} = \hat{\mathbf{H}}(\lambda_0)$. Es gilt

$$\mathbf{E}[\hat{\mathbf{H}}_{\mathbf{M}}|\mathcal{F}] = \hat{\mathbf{F}}_{\mathbf{M}}$$

da $\mathbf{E}[1_{\{k \in r\}}|\mathcal{F}] = \theta_k 1_{\{k \in s\}}$ fast sicher ist und damit der Erwartungswert $\mathbf{E}[\hat{\mathbf{Q}}(\lambda_0)|\mathcal{F}]$ fast sicher Null ist.

Wenn $(s, r) \in \Omega$ eine Realisierung nach einer Stichprobenziehung ist, dann ist $\hat{\lambda}(s, r) \in A$ eine Maximum-Likelihood-Lösung, falls $\mathbf{u}(s, r)(\hat{\lambda}(s, r)) = 0$ gilt und $\hat{\mathbf{H}}(s, r)$ für alle $\lambda \in A$ positiv definit ist. Sei $\hat{\lambda} : \Omega \rightarrow A$ wie folgt definiert. $\hat{\lambda}(s, r)$ ist eine Lösung von $\mathbf{u}(s, r)(\hat{\lambda}(s, r)) = 0$, falls sie existiert und falls keine Lösung existiert, wird $\hat{\lambda}(s, r)$ auf eine beliebigen Konstante $\lambda_A \in A$ gesetzt. Mit dieser Definition gilt mit dem Mittelwertsatz

$$\mathbf{u}(\hat{\lambda}) - \mathbf{u}_{\mathbf{M}} = - \int_0^1 \hat{\mathbf{H}}(\hat{\lambda}_P(a)) da (\hat{\lambda} - \lambda_0)$$

wobei $\hat{\lambda}_P(a) = \lambda_0 + a(\hat{\lambda} - \lambda_0)$ ist. Falls $\mathbf{u}(\lambda) = 0$ fast sicher eine Lösung hat und das Integral auf der rechten Seite asymptotisch durch $\hat{\mathbf{F}}_{\mathbf{M}}$ ersetzt werden kann, dann lässt sich $\hat{\lambda} - \lambda_0$ durch $\hat{\mathbf{F}}_{\mathbf{M}}^{-1} \mathbf{u}_{\mathbf{M}}$ approximieren. Die Invertierbarkeit von $\hat{\mathbf{F}}_{\mathbf{M}}$ sei hierbei vorausgesetzt. Da $\mathbf{E}[\hat{\mathbf{F}}_{\mathbf{M}}^{-1} \mathbf{u}_{\mathbf{M}}|\mathcal{F}] = 0$ und $\mathbf{V}[\hat{\mathbf{F}}_{\mathbf{M}}^{-1} \mathbf{u}_{\mathbf{M}}|\mathcal{F}] = \hat{\mathbf{F}}_{\mathbf{M}}^{-1}$ gilt, lassen sich dann diese beide Eigenschaften auf $\hat{\lambda} - \lambda_0$ übertragen.

Sei $S_{\text{ML}} = \{\exists \lambda \in A : \mathbf{u}(\lambda) = 0\}$ das Ereignis, dass die Gleichung $\mathbf{u}(\lambda) = 0$ eine Lösung hat. Im ersten Schritt soll gezeigt werden, dass unter den Annahmen [A1], [A2] und den Annahmen

[R1] Für die Designgewichte gilt $NM_L/n \leq w_k \leq NM_U/n$ für Konstanten $M_L > 0$ und $M_U > 0$.

[R2] Für beliebige Transformationen $f : B \rightarrow \mathbb{R}^d$ sind die Eigenwerte von $N^{-1} \sum_U 1_{\{k \in s\}} f(\mathbf{x}_k) f(\mathbf{x}_k)'$ für jede Realisierung durch eine Konstante $M_E > 0$ gleichmäßig nach unten beschränkt und $\lim N^{-1} \sum_U f(\mathbf{x}_k) f(\mathbf{x}_k)'$ ist positiv definit.²²

asymptotisch eine Lösung mit Wahrscheinlichkeit 1 existiert, d.h. $P(S_{\text{ML}}) \rightarrow 1$. Hierbei garantieren die Annahmen [R1] und [R2], dass die Fisher-Information invertierbar ist und $n^{-1} \hat{\mathbf{F}}_{\mathbf{M}}$ das gleiche

²²Die Invertierbarkeit von $N^{-1} \sum_U 1_{\{k \in s\}} f(\mathbf{x}_k) f(\mathbf{x}_k)'$ reicht aus, falls man nur die Konvergenz in Wahrscheinlichkeit betrachtet.

asymptotische Verhalten wie eine gewichtete Version der Fisher-Information hat. Außerdem stellen diese beiden Annahmen sicher, dass die Fisher-Information $\hat{\mathbf{F}}$ für alle $\lambda \in A$ positiv definit ist. Somit ist die Lösung von $\mathbf{u}(s, r)(\lambda) = 0$ für $(s, r) \in S_{\text{ML}}$ sogar eindeutig, wenn $f_L = \text{id}$ benutzt wird.

Sei nun eine gewichtete Version von $\hat{\mathbf{F}}$ durch

$$\hat{\mathbf{F}}_w(\lambda) = \sum_U 1_{\{k \in S\}} w_k f_{N_1}(\lambda, \mathbf{x}_k) \mathbf{h}_1(\lambda, \mathbf{x}_k) \mathbf{h}_1(\lambda, \mathbf{x}_k)' \quad (4.6)$$

gegeben. Diese hat den Erwartungswert

$$\mathbf{F}(\lambda) = \sum_U f_{N_1}(\lambda, \mathbf{x}_k) \mathbf{h}_1(\lambda, \mathbf{x}_k) \mathbf{h}_1(\lambda, \mathbf{x}_k)' \quad (4.7)$$

Sei $\lambda_{\min}(\cdot)$ der kleinste Eigenwert einer Matrix. [R2] garantiert, dass sowohl $\hat{\mathbf{F}}_w(\lambda_0)$ als auch $\hat{\mathbf{F}}_M$ und $\mathbf{F}_M = \mathbf{F}(\lambda_0)$ positiv definit sind. Da $|\lambda_{\min}(\hat{\mathbf{F}}_w(\lambda_0)) - \lambda_{\min}(\mathbf{F}_M)| \leq \|\hat{\mathbf{F}}_w(\lambda_0) - \mathbf{F}_M\|$ gilt, konvergiert $\lambda_{\min}(N^{-1}\hat{\mathbf{F}}_w(\lambda_0))$ mit Annahme [A2] in Design-Wahrscheinlichkeit gegen $\lambda_{\min}(N^{-1}\mathbf{F}_M)$. Der Grenzwert von $\lambda_{\min}(N^{-1}\mathbf{F}_M)$ ist positiv nach Annahme [R2]. Damit konvergiert $\lambda_{\min}(N^{-1}\hat{\mathbf{F}}_w(\lambda_0))^{-1}$ gegen $\lambda_{\min}(N^{-1}\mathbf{F}_M)^{-1}$ und ist somit $O_p(1)$. Mit [R1] ist

$$\lambda_{\min}(n^{-1}\hat{\mathbf{F}}_M)^{-1} \leq M_U \lambda_{\min}(N^{-1}\hat{\mathbf{F}}_w(\lambda_0))^{-1} = O_p(1)$$

Also ist $\hat{\mathbf{F}}_M^{-1} = O_p(n^{-1})$.

Die Funktion $N^{-1}\hat{\mathbf{F}}_w$ konvergiert stochastisch gleichmäßig gegen die Funktion $N^{-1}\mathbf{F}$ auf jeder kompakten Menge in A . Denn mit Annahme [A2] konvergiert $N^{-1}\hat{\mathbf{F}}_w$ punktweise in Design-Wahrscheinlichkeit gegen $N^{-1}\mathbf{F}$ und mit Annahme [A1] ist $N^{-1}\hat{\mathbf{F}}_w$ stochastisch gleichgradig und $N^{-1}\mathbf{F}$ gleichgradig stetig auf jeder kompakten Menge in A .

Sowohl Kim und Kim (2007) als auch Kim und Riddles (2012) gehen direkt von der Annahme aus, dass $P(S_{\text{ML}}) = 1$ gilt.

4.2 Existenz der Maximum-Likelihood-Lösung

Im Folgenden wird gezeigt, dass es eine Maximum-Likelihood-Lösung für die Parameter des Response-Modells existiert. Die Existenz ist nicht offensichtlich.

Um die Existenz einer Lösung zu zeigen, wird der Fixpunktsatz von Banach verwendet. Bei diesem Ansatz hat eine kontrahierende Abbildung $\mathbf{Z}: A \rightarrow \mathbb{R}^d$, die eine kompakte Menge $A \subset \mathbb{R}^d$ in sich selbst abbildet genau einen Fixpunkt. Zunächst wird eine Folge von kompakten Mengen gebildet. Sei hierzu die folgende Metrik für $\lambda, \tilde{\lambda} \in A$ definiert:

$$d(\lambda, \tilde{\lambda}) = \frac{1}{M_U} \left\| \sqrt{M_L n N^{-1} \mathbf{F}_M^{1/2}} (\lambda - \tilde{\lambda}) \right\| \quad (4.8)$$

$\mathbf{F}_M^{1/2}$ ist die Quadratwurzel von \mathbf{F}_M , und M_L und M_U sind die Konstanten aus Annahme [R1]. Die Metrik ist so definiert, dass der Abstand zwischen $\lambda \in A$ und $\tilde{\lambda} \in A$ mit der erwarteten Stichprobengröße

wächst.²³ Bildet man mit dieser Metrik die Umgebungen

$$A_n(\lambda_0, \delta) = \{\lambda \in A : d(\lambda, \lambda_0) \leq \delta\}$$

für beliebig gewähltes $\delta > 0$, dann werden diese mit wachsender Stichprobengröße immer kleiner. Das heißt die Umgebungen bilden eine monoton fallende Folge von Mengen, wobei $\bigcap_n A_n(\lambda_0, \delta) = \{\lambda_0\}$ gilt. Die Umgebungen $A_n(\lambda_0, \delta)$ schrumpfen mit $O(n^{1/2})$.

Sei nun $\mathbf{Z}: A \rightarrow \mathbb{R}^d$ definiert durch

$$\mathbf{Z}(\lambda) = \lambda + NM_U(n\mathbf{F}_M)^{-1}\mathbf{u}(\lambda) \quad (4.9)$$

mit erster Ableitung $\mathbf{Z}_1(\lambda) = \mathbf{I} - NM_U(n\mathbf{F}_M)^{-1}\hat{\mathbf{H}}(\lambda)$ wobei \mathbf{I} die Einheitsmatrix ist. Falls $d(\mathbf{Z}(\lambda), \lambda_0) \leq \delta$ für alle $\lambda \in A_n(\lambda_0, \delta)$ fast sicher gilt, dann wird $A_n(\lambda_0, \delta)$ fast sicher auf sich selbst abgebildet. Falls es einen Fixpunkt gibt, hat nach Definition von \mathbf{Z} die Score-Funktion \mathbf{u} fast sicher eine Nullstelle. Dieser Sachverhalt wird im Folgenden in Wahrscheinlichkeit gezeigt.

Mit der oben definierten Metrik ist

$$d(\mathbf{Z}(\lambda), \lambda_0) \leq d(\mathbf{Z}(\lambda), \mathbf{Z}(\lambda_0)) + d(\mathbf{Z}(\lambda_0), \lambda_0) \quad (4.10)$$

Für den zweiten Term auf der rechten Seite der Ungleichung gilt

$$d(\mathbf{Z}(\lambda_0), \lambda_0) = \left\| \sqrt{M_L N n^{-1} \mathbf{F}_M^{-1/2} \mathbf{u}_M} \right\|$$

Wie bei Fahrmeir und Kaufmann (1985) soll zunächst gezeigt werden, dass der Erwartungswert von $d(\mathbf{Z}(\lambda_0), \lambda_0)^2$ durch eine Konstante beschränkt ist. Mit der Annahme [R1] gilt für jeden Vektor $\nu \in \mathbb{R}^d$ der Länge 1, dass

$$M_L \frac{N}{n} \nu' \mathbf{F}_M^{-1/2} \hat{\mathbf{F}}_M \mathbf{F}_M^{-1/2} \nu \leq \nu' \mathbf{F}_M^{-1/2} \hat{\mathbf{F}}_W \mathbf{F}_M^{-1/2} \nu$$

gilt. Hierbei ist der Erwartungswert der rechten Seite 1, da $E[\hat{\mathbf{F}}_W] = \mathbf{F}_M$ gilt. Das bedeutet insbesondere, dass alle Diagonaleinträge der Matrix

$$M_L \frac{N}{n} E \left[\mathbf{F}_M^{-1/2} \hat{\mathbf{F}}_M \mathbf{F}_M^{-1/2} \right] = M_L \frac{N}{n} E \left[\mathbf{F}_M^{-1/2} E[\mathbf{u}_M \mathbf{u}'_M | \mathcal{F}] \mathbf{F}_M^{-1/2} \right]$$

kleiner als 1 sind. Folglich ist die Spur dieser Matrix kleiner als die Dimension d . Die Spur ist aber $E[d(\mathbf{Z}(\lambda_0), \lambda_0)^2]$. Somit existiert für vorgegebenes $\varepsilon > 0$ und vorgegebenes $c \in (0, 1)$ ein $\delta > 0$, so dass mit der Markov-Ungleichung

$$P(\{d(\mathbf{Z}(\lambda_0), \lambda_0) > (1 - c)\delta\}) \leq (1 - c)^{-1} d/\delta < \varepsilon/2 \quad (4.11)$$

gilt.

²³ \mathbf{F}_M ist positiv definit und konvergiert mit Annahme [R2] gegen eine positiv definite Matrix. Daher ist der kleinste Eigenwert $\lambda_{\min}(N^{-1}\mathbf{F}_M)$ nach unten durch eine positive Konstante beschränkt. Da außerdem

$$\lambda_{\min}(\mathbf{F}_M)^{1/2} \|\lambda - \tilde{\lambda}\| \leq \|\mathbf{F}_M^{1/2}(\lambda - \tilde{\lambda})\|$$

gilt, wächst $d(\lambda, \tilde{\lambda})$ mit $n^{1/2}$.

Als nächstes wird $d(\mathbf{Z}(\lambda), \mathbf{Z}(\lambda_0))$ betrachtet. Mit dem Mittelwertsatz gilt

$$\mathbf{Z}(\lambda) - \mathbf{Z}(\lambda_0) = \int_0^1 \mathbf{Z}_1(\lambda_P(a)) da (\lambda - \lambda_0) = \int_0^1 \mathbf{I} - NM_U(n\mathbf{F}_M)^{-1} \hat{\mathbf{H}}(\lambda_P(a)) da (\lambda - \lambda_0)$$

wobei $\lambda_P(a) = \lambda_0 + a(\lambda - \lambda_0)$ ist. Daher ist

$$\begin{aligned} d(\mathbf{Z}(\lambda), \mathbf{Z}(\lambda_0)) &\leq \int_0^1 \left\| \mathbf{I} - NM_U n^{-1} \mathbf{F}_M^{-1/2} \hat{\mathbf{H}}(\lambda_P(a)) \mathbf{F}_M^{-1/2} \right\| da d(\lambda, \lambda_0) \\ &\leq \int_0^1 \left\| \mathbf{I} - NM_U n^{-1} \mathbf{F}_M^{-1/2} \hat{\mathbf{F}}(\lambda_P(a)) \mathbf{F}_M^{-1/2} \right\| da d(\lambda, \lambda_0) + \\ &\quad \int_0^1 \left\| NM_U n^{-1} \mathbf{F}_M^{-1/2} \hat{\mathbf{Q}}(\lambda_P(a)) \mathbf{F}_M^{-1/2} \right\| da d(\lambda, \lambda_0) \\ &\leq \int_0^1 \left\| \mathbf{I} - \mathbf{F}_M^{-1/2} \hat{\mathbf{F}}_W(\lambda_P(a)) \mathbf{F}_M^{-1/2} \right\| da d(\lambda, \lambda_0) + \\ &\quad \left\| M_U (N^{-1} \mathbf{F}_M)^{-1} \right\| \int_0^1 \left\| n^{-1} \hat{\mathbf{Q}}(\lambda_P(a)) \right\| da d(\lambda, \lambda_0) \end{aligned}$$

mit Annahme [R1]. Außerdem ist

$$\int_0^1 \left\| \mathbf{I} - \mathbf{F}_M^{-1/2} \hat{\mathbf{F}}_W(\lambda_P(a)) \mathbf{F}_M^{-1/2} \right\| da \leq \left\| (N^{-1} \mathbf{F}_M)^{-1} \right\| \int_0^1 \left\| N^{-1} (\hat{\mathbf{F}}_W(\lambda_P(a)) - \mathbf{F}_M) \right\| da$$

Für alle $\lambda \in A_n(\lambda_0, \delta)$ gilt damit

$$d(\mathbf{Z}(\lambda), \mathbf{Z}(\lambda_0)) \leq \delta M \left(\sup_{A_n(\lambda_0, \delta)} \left\| N^{-1} (\hat{\mathbf{F}}_W(\tilde{\lambda}) - \mathbf{F}_M) \right\| + \sup_{A_n(\lambda_0, \delta)} \left\| n^{-1} \hat{\mathbf{Q}}(\tilde{\lambda}) \right\| \right)$$

wobei $M > 0$ so gewählt wird, so dass sowohl $\left\| (N^{-1} \mathbf{F}_M)^{-1} \right\|$ als auch $\left\| M_U (N^{-1} \mathbf{F}_M)^{-1} \right\|$ durch M für alle n beschränkt wird. Dies ist durch Annahme [R2] möglich. Der erste Term auf der rechten Seite konvergiert in Design-Wahrscheinlichkeit gegen Null, da $N^{-1} \hat{\mathbf{F}}_W$ stochastisch gleichmäßig gegen die Funktion $N^{-1} \mathbf{F}$ konvergiert. Die Konvergenz des zweiten Terms wird im Anhang A.2 gezeigt. Folglich gibt es zu obigen $\delta > 0$ und gewählten $\varepsilon > 0$ und $c \in (0, 1)$ eine Stichprobengröße $n \in \mathbb{N}$, so dass

$$P(\{\exists \lambda \in A_n(\lambda_0, \delta) : d(\mathbf{Z}(\lambda), \mathbf{Z}(\lambda_0)) > c\delta\}) < \varepsilon/2 \quad (4.12)$$

gilt. Mit dem Fixpunktsatz von Banach²⁴ gibt es für $(s, r) \in \Omega$ mit $d(\mathbf{Z}(s, r)(\lambda), \mathbf{Z}(s, r)(\lambda_0)) \leq c\delta$ ein eindeutiges $\tilde{\lambda} \in A_n(\lambda_0, \delta)$, so dass $\mathbf{Z}(s, r)(\tilde{\lambda}) = \tilde{\lambda}$ gilt.

Mit (4.10), (4.11) und (4.12) gibt es zu $\varepsilon > 0$ ein $\delta > 0$ und eine Stichprobengröße $n \in \mathbb{N}$, so dass

$$P(\{\exists \lambda \in A_n(\lambda_0, \delta) : d(\mathbf{Z}(\lambda), \lambda_0) > \delta\}) < \varepsilon$$

ist. Das heißt, die Umgebung $A_n(\lambda_0, \delta)$ wird mit einer Wahrscheinlichkeit größer $1 - \varepsilon$ auf sich selbst abgebildet, falls die Stichprobe hinreichend groß ist. Da es mit dem Fixpunktsatz von Banach einen Fixpunkt für $(s, r) \in \Omega$ mit $d(\mathbf{Z}(s, r)(\lambda), \mathbf{Z}(s, r)(\lambda_0)) \leq c\delta$ gibt, ist mit der Definition der Metrik und der Definition von \mathbf{Z}

$$P(\{\exists \lambda \in A_n(\lambda_0, \delta) : \mathbf{u}(\lambda) = 0\}) \geq 1 - \varepsilon$$

²⁴Die kontrahierende Abbildung ist unabhängig von der Wahl von λ_0 .

Definiere $\hat{\lambda} : \Omega \rightarrow A$ als Lösung von $\mathbf{u}(s, r)(\lambda) = 0$ für alle $(s, r) \in \Omega$, falls sie existiert. Sonst wähle $\hat{\lambda}(s, r) = \lambda_A$ wobei λ_A auf dem Rand von A liegt. Da $P(\{\mathbf{u}(\hat{\lambda}) = 0\}) \geq P(\{\exists \lambda \in A_n(\lambda_0, \delta) : \mathbf{u}(\lambda) = 0\})$ gilt, folgt die Existenz einer Lösung mit $P(\{\mathbf{u}(\hat{\lambda}) = 0\}) \rightarrow 1$.

Das Ereignis $\{\hat{\lambda} \in A_n(\lambda_0, \delta)\}$ ist nur zum Ereignis $\{\exists \lambda \in A_n(\lambda_0, \delta) : \mathbf{u}(\lambda) = 0\}$ äquivalent falls $\hat{\mathbf{H}}$ für alle $\lambda \in A$ fast sicher positiv definit ist. Dies ist für $f_L = \text{id}$ der Fall. Sonst ist die Lösung nicht eindeutig. Es folgt

$$P(\{\sqrt{n}\|\hat{\lambda} - \lambda_0\| > M\delta\}) = P(\{\hat{\lambda} \notin A_n(\lambda_0, \delta)\}) < \varepsilon$$

Somit existiert eine Maximum-Likelihood-Lösung unter den schwachen Annahmen [A1], [A2], [R1] und [R2]. [R1] ist eine Einschränkung, dass nicht jedes Stichprobendesign gewählt werden kann.

4.3 Approximation der Maximum-Likelihood-Lösung

Im letzten Abschnitt wurde gezeigt, dass eine Lösung der Gleichung $\mathbf{u}(\lambda) = 0$ mit einer Wahrscheinlichkeit von Eins asymptotisch existiert. Es gibt aber für eine endliche Stichprobe immer eine (kleine) Restwahrscheinlichkeit die zur Nicht-Existenz einer Lösung führen kann. Betrachtet man hierzu

$$\mathbf{u}(\hat{\lambda}) - \mathbf{u}_M = - \int_0^1 \hat{\mathbf{H}}(\hat{\lambda}_P(a)) da (\hat{\lambda} - \lambda_0)$$

wobei $\hat{\lambda}_P(a) = \lambda_0 + a(\hat{\lambda} - \lambda_0)$ ist. Dann gilt die Gleichung

$$\mathbf{u}_M = \int_0^1 \hat{\mathbf{H}}(\hat{\lambda}_P(a)) da (\hat{\lambda} - \lambda_0) \tag{4.13}$$

nur dann fast sicher, falls $\hat{\lambda} : \Omega \rightarrow A$ fast sicher eine Lösung ist.²⁵ Im Vergleich zum Kalibrationsansatz muss die folgende Annahme getroffen werden

[R3] Die Gleichung $\mathbf{u}(\lambda) = 0$ hat fast sicher eine Lösung.

Die Gleichung (4.13) lässt sich umformen

$$\begin{aligned} (\hat{\lambda} - \lambda_0) - \hat{\mathbf{F}}_M^{-1} \mathbf{u}_M &= \hat{\mathbf{F}}_M^{-1} \int_0^1 (\hat{\mathbf{F}}_M - \hat{\mathbf{H}}(\hat{\lambda}_P(a))) da (\hat{\lambda} - \lambda_0) \\ &= \int_0^1 \hat{\mathbf{F}}_M^{-1} (\hat{\mathbf{F}}_M - \hat{\mathbf{F}}(\hat{\lambda}_P(a))) da (\hat{\lambda} - \lambda_0) - \int_0^1 \hat{\mathbf{F}}_M^{-1} \hat{\mathbf{Q}}(\hat{\lambda}_P(a)) da (\hat{\lambda} - \lambda_0) \end{aligned}$$

Sei $f_{N2}(\lambda, \mathbf{x}_k) = b_3(f_L(\lambda' \mathbf{x}_k))$ und $f_S : B \rightarrow \mathbb{R}^+$ eine Funktion mit

$$\|f_{N2}(\lambda, \mathbf{x}_k) f_{L1}(\lambda' \mathbf{x}_k)^3 \mathbf{x}_k + 2f_{N1}(\lambda, \mathbf{x}_k) f_{L1}(\lambda' \mathbf{x}_k) f_{L2}(\lambda' \mathbf{x}_k) \mathbf{x}_k\| \text{Spur} |\mathbf{x}_k \mathbf{x}_k'| \leq f_S(\mathbf{x}_k)$$

für alle $\lambda \in A_K$, $A_K \subset A$ kompakt. Mit Annahme [R1] und wiederum dem Mittelwertsatz ist dann für $\lambda, \tilde{\lambda} \in A_K$

$$\|\hat{\mathbf{F}}_M^{-1} (\hat{\mathbf{F}}(\lambda) - \hat{\mathbf{F}}(\tilde{\lambda}))\| \leq \left\| (n^{-1} \hat{\mathbf{F}}_M)^{-1} \right\| \frac{1}{NM_L} \sum_U 1_{\{k \in s\}} w_k f_S(\mathbf{x}_k) \|\lambda - \tilde{\lambda}\| \tag{4.14}$$

²⁵Bei Fahrmeir und Kaufmann (1985) Theorem 3 wird argumentiert, dass die Wahrscheinlichkeit für eine Lösung nahe 1 ist.

Sei $\varepsilon > 0$ und $\delta > 0$ beliebig, so dass $P(\{\hat{\lambda} \notin A_n(\lambda_0, \delta)\}) < \varepsilon/3$ für die Umgebungen $A_n(\lambda_0, \delta) = \{\lambda \in A : \sqrt{n}\|\lambda - \lambda_0\| \leq \delta\}$ gilt. Dann gibt es mit den Annahmen [A2], [R1] und [R2] ein $M > 0$, so dass sowohl

$$P\left(\left\{\delta \sup_{A_n(\lambda_0, \delta)} \sqrt{n} \left\| \hat{\mathbf{F}}_M^{-1}(\hat{\mathbf{F}}(\tilde{\lambda}) - \hat{\mathbf{F}}(\lambda_0)) \right\| > M/2 \right\}\right) < \varepsilon/3$$

als auch (siehe Anhang A.2)

$$P\left(\left\{\delta \sup_{A_n(\lambda_0, \delta)} \sqrt{n} \left\| \hat{\mathbf{F}}_M^{-1} \hat{\mathbf{Q}}(\tilde{\lambda}) \right\| > M/2 \right\}\right) < \varepsilon/3$$

ist. Damit ist

$$\begin{aligned} P\left(\left\{\|n((\hat{\lambda} - \lambda_0) - \hat{\mathbf{F}}_M^{-1} \mathbf{u}_M)\| > M\right\}\right) &\leq P\left(\left\{\|n((\hat{\lambda} - \lambda_0) - \hat{\mathbf{F}}_M^{-1} \mathbf{u}_M)\| > M, \hat{\lambda} \in A_n(\lambda_0, \delta)\right\}\right) + \\ &\quad P\left(\left\{\|n((\hat{\lambda} - \lambda_0) - \hat{\mathbf{F}}_M^{-1} \mathbf{u}_M)\| > M, \hat{\lambda} \notin A_n(\lambda_0, \delta)\right\}\right) \\ &\leq P\left(\left\{\delta \sup_{A_n(\lambda_0, \delta)} \sqrt{n} \left\| \hat{\mathbf{F}}_M^{-1}(\hat{\mathbf{F}}(\tilde{\lambda}) - \hat{\mathbf{F}}(\lambda_0)) \right\| > M/2\right\}\right) + \\ &\quad P\left(\left\{\delta \sup_{A_n(\lambda_0, \delta)} \sqrt{n} \left\| \hat{\mathbf{F}}_M^{-1} \hat{\mathbf{Q}}(\tilde{\lambda}) \right\| > M/2\right\}\right) + \\ &\quad P\left(\left\{\hat{\lambda} \notin A_n(\lambda_0, \delta)\right\}\right) \\ &\leq \varepsilon \end{aligned}$$

und somit

$$\hat{\lambda} - \lambda_0 = \hat{\mathbf{F}}_M^{-1} \mathbf{u}_M + O_p(n^{-1}) \quad (4.15)$$

Um die Aussage (4.15) zu verschärfen, muss zusätzlich zur Annahme [R3] auch noch die Annahme, dass der Parameter-Raum kompakt ist, getroffen werden ([R4] Der Parameter-Raum ist kompakt). Damit ist der kleinste Eigenwert durch eine Konstant $M_E > 0$ beschränkt:

$$\lambda_{\min} \left(n^{-1} \int_0^1 \hat{\mathbf{F}}(\hat{\lambda}_P(a)) da \right)^{-1} < M_E$$

Mit der Gleichung (4.13) ist

$$\left(n^{-1} \int_0^1 \hat{\mathbf{F}}(\hat{\lambda}_P(a)) da \right)^{-1} \left(\frac{1}{\sqrt{n}} \mathbf{u}_M - \frac{1}{\sqrt{n}} \int_0^1 \hat{\mathbf{Q}}(\hat{\lambda}_P(a)) da (\hat{\lambda} - \lambda_0) \right) = \sqrt{n}(\hat{\lambda} - \lambda_0) \quad (4.16)$$

Mit der Tschebyscheff-Ungleichung ist leicht zu sehen, dass $n^{-1} \mathbf{u}_M$ in Wahrscheinlichkeit gegen Null konvergiert. Die bedingte Varianz von $n^{-1/2} \mathbf{u}_M$ ist $n^{-1} \hat{\mathbf{F}}_M$, die mit Annahme [R2] beschränkt ist. Die gewichtete Score-Funktion

$$\mathbf{u}_W(\lambda) = \sum_U 1_{\{k \in r\}} w_k \mathbf{x}_k - \sum_U 1_{\{k \in s\}} w_k \mathbf{x}_k f_N(\lambda, \mathbf{x}_k)$$

hat an der Stelle $\lambda_0 \in A$ den Erwartungswert Null. Mit den Annahmen [A2] und [A3] gilt für die i -te

Zeile und $p \in \mathbb{N}$

$$\mathbb{E} \left[\left(\frac{\sqrt{n}}{N} \sum_U 1_{\{k \in s\}} w_k (1_{\{k \in r\}} - \theta_k) \mathbf{x}_{ik} \right)^{2p} \right] < M$$

für ein $M > 0$, da der Grenzwert existiert. Da mit Annahme [R1] die Ungleichung

$$\frac{1}{n^p} \sum_U 1_{\{k \in s\}} \mathbf{x}_{ik}^{2p} \mathbb{E}[(1_{\{k \in r\}} - \theta_k)^{2p} | \mathcal{F}] \leq \frac{n^p}{(M_L N)^{2p}} \sum_U 1_{\{k \in s\}} (w_k \mathbf{x}_{ik})^{2p} \mathbb{E}[(1_{\{k \in r\}} - \theta_k)^{2p} | \mathcal{F}]$$

für $p \in \mathbb{N}$ gilt, gilt bei Unabhängigkeit²⁶

$$\mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \sum_U 1_{\{k \in s\}} (1_{\{k \in r\}} - \theta_k) \mathbf{x}_{ik} \right)^{2p} \right] \leq \mathbb{E} \left[\left(\frac{\sqrt{n}}{M_L N} \sum_U 1_{\{k \in s\}} w_k (1_{\{k \in r\}} - \theta_k) \mathbf{x}_{ik} \right)^{2p} \right]$$

Mit der Zerlegung in A.2 und einer analogen Vorgehensweise, ist

$$\mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \int_0^1 \hat{\mathbf{Q}}(\hat{\lambda}_P(a)) da \right)^{2p} \right] < M$$

für ein $M > 0$. Da $\hat{\lambda}$ mit Annahme [R4] in einer kompakten Menge liegt, ist die rechte Seite von (4.16) im p -ten Mittel beschränkt. Damit ist der Erwartungswert von $n^p \|\hat{\lambda} - \lambda_0\|^{2p}$ beschränkt und

$$\mathbb{E} \left[\left(\frac{1}{\sqrt{n}} \int_0^1 \hat{\mathbf{Q}}(\hat{\lambda}_P(a)) da (\hat{\lambda} - \lambda_0) \right)^{2p} \right] = O(n^{-p})$$

Mit (4.14) ist dann auch der Erwartungswert von $n^p \|\hat{\lambda} - \lambda_0\| - \hat{\mathbf{F}}_M^{-1} \mathbf{u}_M\|^p$ beschränkt.

4.4 Varianz des PSA-Schätzers

Im folgenden wird die Varianz des PSA-Schätzers hergeleitet. Seien f_{M1} und f_{M2} die ersten und zweiten Ableitungen von f_M . Diese können mit $f_M(\lambda_0, \mathbf{x}_k)$ normiert werden: $f_{K1}(\lambda, \mathbf{x}_k) = f_{M1}(\lambda, \mathbf{x}_k) / f_M(\lambda_0, \mathbf{x}_k)$ und $f_{K2}(\lambda, \mathbf{x}_k) = f_{M2}(\lambda, \mathbf{x}_k) / f_M(\lambda_0, \mathbf{x}_k)$. Außerdem sei das Wegintegral

$$F_{K2}(\hat{\lambda}_P, \cdot) = \int_0^1 (1-a) f_{K2}(\hat{\lambda}_P(a), \cdot) da$$

mit $\hat{\lambda}_P(a) = \lambda_0 + a(\hat{\lambda} - \lambda_0)$ gegeben. Mit

$$\begin{aligned} \hat{\mathbf{D}}_M &= N^{-1} \sum_U 1_{\{k \in r\}} w_{Mk} y_k f_{K1}(\lambda_0, \mathbf{x}_k) \text{ und} \\ \hat{\mathbf{R}}_M(\hat{\lambda}_P) &= N^{-1} \sum_U 1_{\{k \in r\}} w_{Mk} y_k F_{K2}(\hat{\lambda}_P, \mathbf{x}_k) \end{aligned}$$

lässt sich (2.10) darstellen als

$$\hat{Y}_M = \hat{Y}_M + \hat{\mathbf{D}}_M' (\hat{\lambda} - \lambda_0) + \frac{1}{2} (\hat{\lambda} - \lambda_0)' \hat{\mathbf{R}}_M(\hat{\lambda}_P) (\hat{\lambda} - \lambda_0)$$

²⁶Terme, die $1_{\{k \in r\}} - \theta_k$ nur einmal enthalten, sind wegen der Unabhängigkeit Null.

Dies Gleichung lässt sich umformen zu

$$\hat{Y}_{\hat{M}} - Y = (\hat{Y}_M - Y) + \mathbf{D}'_M(\hat{\lambda} - \lambda_0) + (\hat{\mathbf{D}}_M - \mathbf{D}_M)'(\hat{\lambda} - \lambda_0) + \frac{1}{2}(\hat{\lambda} - \lambda_0)' \hat{\mathbf{R}}_M(\hat{\lambda}_P)(\hat{\lambda} - \lambda_0)$$

mit $\mathbf{D}_M = N^{-1} \sum_U y_k f_{K1}(\lambda_0, \mathbf{x}_k)$. Mit den Annahme [A1] und [A2] ist $\hat{\mathbf{R}}_M(\hat{\lambda}_P)$ in Wahrscheinlichkeit beschränkt, siehe Abschnitt 2.4. Außerdem ist $\hat{\mathbf{D}}_M - \mathbf{D}_M = O_p(n^{-1/2})$ mit der Annahme [A2]. Damit gilt

$$\hat{Y}_{\hat{M}} - Y = (\hat{Y}_M - Y) + \mathbf{D}'_M(\hat{\lambda} - \lambda_0) + O_p(n^{-1})$$

Und mit (4.15) ist

$$\hat{Y}_{\hat{M}} - Y = (\hat{Y}_M - Y) + \mathbf{D}'_M \hat{\mathbf{F}}_M^{-1} \mathbf{u}_M + O_p(n^{-1})$$

Mit den Ausführungen des letzten Abschnitts gilt:

Satz (PSA-Schätzer bei Nonresponse). Mit den Response-Wahrscheinlichkeiten $\theta_k = b_1(f_L(\lambda' \mathbf{x}_k))$ ist bei Quasi-Randomization und den Annahmen [A1], [A2], [A3] und [R1]-[R4] der Bias von $\hat{Y}_{\hat{M}}$

$$E[\hat{Y}_{\hat{M}} - Y] = O(n^{-1}) \quad (4.17)$$

und unter Beachtung von $E[\hat{Y}_M \mathbf{u}_M] = -\mathbf{D}_M$ gilt für die Varianz

$$V[\hat{Y}_{\hat{M}} - Y] = V[\hat{Y}_M - Y] - E[\mathbf{D}'_M \hat{\mathbf{F}}_M^{-1} \mathbf{D}_M] + O(n^{-2})$$

□

Mit (2.8) ist die Varianz

$$V[\hat{Y}_{\hat{M}}] = \frac{1}{N^2} \sum \sum_U \Delta_{kl} \frac{y_k y_l}{\pi_k \pi_l} + \frac{1}{N^2} E[\sum_U 1_{\{k \in s\}} \theta_k (1 - \theta_k) (w_{Mk} y_k)^2] - E[\mathbf{D}'_M \hat{\mathbf{F}}_M^{-1} \mathbf{D}_M] + O(n^{-2}) \quad (4.18)$$

Mit (2.9) ist der Varianzschätzer für $f_L = \text{id}$

$$\hat{V}_M = \frac{1}{N^2} \sum \sum_U 1_{\{k, l \in r\}} \frac{\Delta_{kl}}{\pi_{kl}} (\hat{w}_{Mk} y_k) (\hat{w}_{Ml} y_l) + \frac{1}{N^2} \sum_U 1_{\{k \in r\}} \pi_k (1 - \hat{\theta}_k) (\hat{w}_{Mk} y_k)^2 - \hat{\mathbf{D}}'_M \hat{\mathbf{F}}_M^{-1} \hat{\mathbf{D}}_M \quad (4.19)$$

wobei $\hat{w}_{Mk} = w_k \{1 + \exp(-\hat{\lambda}' \mathbf{x}_k)\}$, $\hat{\theta}_k = \{1 + \exp(-\hat{\lambda}' \mathbf{x}_k)\}^{-1}$ und

$$\begin{aligned} \hat{\mathbf{D}}_M &= -N^{-1} \sum_U 1_{\{k \in r\}} w_k y_k \mathbf{x}_k (1 - \hat{\theta}_k) \\ \hat{\mathbf{F}}_M &= \sum_U 1_{\{k \in s\}} \hat{\theta}_k (1 - \hat{\theta}_k) \mathbf{x}_k \mathbf{x}'_k \end{aligned}$$

Man beachte hierbei, dass $\hat{\mathbf{F}}_M$ auf Basis der gesamten Stichprobe berechnet wird.

4.5 Kalibration des PSA-Schätzers

Es soll im Folgenden der PSA-Schätzer zusätzlich kalibriert werden. Für jede Einheit $k \in U$ seien zwei unterschiedliche Mengen an Hilfsmerkmalen \mathbf{x}_{Ak} und \mathbf{x}_{Bk} gegeben. Für die A Merkmale ist auch der Mittelwert $\mathbf{X}_A = N^{-1} \sum_U \mathbf{x}_{Ak}$ bekannt. Mit den B Merkmalen werden zuerst die Gewichte des PSA-Schätzers $\hat{w}_{Mk} = w_k \{1 + \exp(-\hat{\lambda}' \mathbf{x}_{Bk})\}$ berechnet. Ziel ist es, die Kalibrationsgewichte $\hat{w}_{Ck}(s, r)$ auf Basis der Auskunftspersonen so zu bestimmen, dass

$$\sum_U 1_{\{k \in r\}} (\hat{w}_{Ck} - \hat{w}_{Mk})^2 / 2 \hat{w}_{Mk}$$

unter der Nebenbedingung

$$\frac{1}{N} \sum_U 1_{\{k \in r\}} \hat{w}_{Ck} \mathbf{x}_{Ak} = \mathbf{X}_A$$

minimal wird. Mit diesen Gewichten ist der kalibrierte PSA-Schätzer gegeben durch

$$\hat{Y}_{\hat{C}} = \frac{1}{N} \sum_U 1_{\{k \in r\}} \hat{w}_{Ck} y_k \quad (4.20)$$

Die Lösung des obigen Minimierungsproblems mit dem Lagrange-Ansatz ergibt:

$$\hat{w}_{Ck}(s, r) = \hat{w}_{Mk}(s, r) (1 + \beta(s, r)' \mathbf{x}_{Ak})$$

für $k \in r$, wobei β der Lagrange-Multiplikator ist. Einsetzen von $\hat{w}_{Ck}(s, r)$ in die Nebenbedingung ergibt

$$\beta = (\sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_{Ak} \mathbf{x}'_{Ak})^{-1} (N \mathbf{X}_A - \sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_{Ak})$$

Hierbei muss gewährleistet sein, dass $\sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_{Ak} \mathbf{x}'_{Ak}$ invertierbar ist (Annahme [R2]). Durch das Einsetzen von γ in die Gleichung von $\hat{w}_{Ck}(s, r)$ sind die Gewichte $\hat{w}_{Ck}(s, r)$ durch

$$\hat{w}_{Ck} = \hat{w}_{Mk} + (N \mathbf{X}_A - \sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_{Ak})' (\sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_{Ak} \mathbf{x}'_{Ak})^{-1} \hat{w}_{Mk} \mathbf{x}_{Ak} \quad (4.21)$$

gegeben. Der Kalibrationsschätzer in (4.20) lässt sich nun mit den Gewichten (4.21) berechnen. Der Kalibrationsschätzer lässt sich auch darstellen als

$$\hat{Y}_{\hat{C}} = \hat{Y}_{\hat{M}} + \hat{\mathbf{B}}'_{\hat{M}} (\mathbf{X}_A - \hat{\mathbf{X}}_{\hat{M}}) \quad (4.22)$$

mit

$$\hat{\mathbf{B}}_{\hat{M}} = (\sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_{Ak} \mathbf{x}'_{Ak})^{-1} \sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_{Ak} y_k$$

und dem PSA-Schätzer

$$\hat{\mathbf{X}}_{\hat{M}} = \frac{1}{N} \sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_{Ak}$$

Damit man das asymptotische Verhalten von (4.20) untersucht werden kann, seien noch die fol-

genden Größen gegeben.

$$\begin{aligned}\mathbf{B} &= (\sum_U \mathbf{x}_{Ak} \mathbf{x}'_{Ak})^{-1} \sum_U \mathbf{x}_{Ak} y_k \\ \hat{\mathbf{T}}_{\hat{M}} &= N^{-1} \sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_{Ak} \mathbf{x}'_{Ak} \\ \mathbf{T} &= N^{-1} \sum_U \mathbf{x}_{Ak} \mathbf{x}'_{Ak}\end{aligned}$$

\mathbf{B} ist der Populationswert von $\hat{\mathbf{B}}_{\hat{M}}$. \mathbf{T} ist der Populationswert von $\hat{\mathbf{T}}_{\hat{M}}$. Wegen der Ungleichung $|\lambda_{\min}(\hat{\mathbf{T}}_{\hat{M}}) - \lambda_{\min}(\mathbf{T})| \leq \|\hat{\mathbf{T}}_{\hat{M}} - \mathbf{T}\|$, wobei $\lambda_{\min}(\cdot)$ den kleinsten Eigenwert einer Matrix bezeichnet, und da $\hat{\mathbf{T}}_{\hat{M}}$ ein PSA-Schätzer ist, konvergiert die Differenz $\lambda_{\min}(\hat{\mathbf{T}}_{\hat{M}}) - \lambda_{\min}(\mathbf{T})$ in Wahrscheinlichkeit gegen Null. Mit Annahme [R2] und der Bemerkung zum Continuous Mapping Theorem ist $\lambda_{\min}(\hat{\mathbf{T}}_{\hat{M}})^{-1} = O_p(1)$. Da $\lambda_{\min}(\hat{\mathbf{T}}_{\hat{M}})^{-1}$ der größte Eigenwert von $\hat{\mathbf{T}}_{\hat{M}}^{-1}$ ist, ist $\hat{\mathbf{T}}_{\hat{M}}^{-1} = O_p(1)$. Wegen $\sum_U \mathbf{x}_{Ak} (y_k - \mathbf{x}'_{Ak} \mathbf{B}) = 0$ ist

$$\frac{1}{N} \sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_{Ak} (y_k - \mathbf{x}'_{Ak} \mathbf{B}) = O_p(n^{-1/2})$$

da auch dies ein PSA-Schätzer ist. Somit ist die Differenz

$$\hat{\mathbf{B}}_{\hat{M}} - \mathbf{B} = (\sum_U 1_{\{k \in r\}} \hat{w}_{Mk} \mathbf{x}_{Ak} \mathbf{x}'_{Ak})^{-1} \sum_U 1_{\{k \in s\}} \hat{w}_{Mk} \mathbf{x}_{Ak} (y_k - \mathbf{x}'_{Ak} \mathbf{B})$$

$O_p(n^{-1/2})$. Unter der Annahme eines kompakten Parameter-Raums, Annahme [R4], ist sogar das p -te Mittel von $\sqrt{n}(\hat{\mathbf{B}}_{\hat{M}} - \mathbf{B})$ beschränkt.

Mit (4.22) ist

$$\hat{Y}_{\hat{C}} - Y = \hat{Y}_{\hat{M}} - Y + \mathbf{B}'(\mathbf{X}_A - \hat{\mathbf{X}}_{\hat{M}}) + (\hat{\mathbf{B}}_{\hat{M}} - \mathbf{B})'(\mathbf{X}_A - \hat{\mathbf{X}}_{\hat{M}})$$

Der letzte Term ist $O_p(n^{-1})$. Mit Annahme [R4] ist $E[\hat{Y}_{\hat{C}} - Y] = O(n^{-1})$ und die Varianz ist mit $e_k = y_k - \mathbf{x}'_{Ak} \mathbf{B}$ auf Basis von (4.18)

$$V[\hat{Y}_{\hat{C}}] = \frac{1}{N^2} \sum \sum_U \Delta_{kl} \frac{e_k}{\pi_k} \frac{e_l}{\pi_l} + \frac{1}{N^2} E[\sum_U 1_{\{k \in s\}} \theta_k (1 - \theta_k) (w_{Mk} e_k)^2] - E[\mathbf{D}'_{\hat{M}} \hat{\mathbf{F}}_{\hat{M}}^{-1} \mathbf{D}_{\hat{M}}] + O(n^{-2}) \quad (4.23)$$

wobei $w_{Mk} = w_k \{1 + \exp(-\lambda_0' \mathbf{x}_{Bk})\}$, $\theta_k = \{1 + \exp(-\lambda_0' \mathbf{x}_{Bk})\}^{-1}$ und

$$\begin{aligned}\mathbf{D}_{\hat{M}} &= -\frac{1}{N} \sum_U e_k \mathbf{x}_{Bk} (1 - \theta_k) \\ \hat{\mathbf{F}}_{\hat{M}} &= \sum_U 1_{\{k \in s\}} \theta_k (1 - \theta_k) \mathbf{x}_{Bk} \mathbf{x}'_{Bk}\end{aligned}$$

Mit (4.19) ist der Varianzschätzer

$$\hat{V}_{\hat{C}} = \frac{1}{N^2} \sum \sum_U 1_{\{k, l \in r\}} \frac{\Delta_{kl}}{\pi_{kl}} (\hat{w}_{Mk} \hat{e}_k) (\hat{w}_{Ml} \hat{e}_l) + \frac{1}{N^2} \sum_U 1_{\{k \in r\}} \pi_k (1 - \hat{\theta}_k) (\hat{w}_{Ml} \hat{e}_k)^2 - \hat{\mathbf{D}}'_{\hat{M}} \hat{\mathbf{F}}_{\hat{M}}^{-1} \hat{\mathbf{D}}_{\hat{M}} \quad (4.24)$$

wobei $\hat{e}_k = y_k - \mathbf{x}'_{Ak} \hat{\mathbf{B}}_{\hat{M}}$, $\hat{w}_{Mk} = w_k \{1 + \exp(-\hat{\lambda}' \mathbf{x}_{Bk})\}$, $\hat{\theta}_k = \{1 + \exp(-\hat{\lambda}' \mathbf{x}_{Bk})\}^{-1}$ und

$$\begin{aligned}\hat{\mathbf{D}}_{\hat{M}} &= -N^{-1} \sum_U 1_{\{k \in r\}} w_k \hat{e}_k \mathbf{x}_{Bk} (1 - \hat{\theta}_k) \\ \hat{\mathbf{F}}_{\hat{M}} &= \sum_U 1_{\{k \in s\}} \hat{\theta}_k (1 - \hat{\theta}_k) \mathbf{x}_{Bk} \mathbf{x}'_{Bk}\end{aligned}$$

4.6 Die Varianz bei Abhängigkeiten

Bei den bisherigen Ausführungen wurde angenommen, dass alle Einheiten unabhängig voneinander sind. Allerdings kann nicht davon ausgegangen werden, dass die räumliche Mobilität zwischen den Einheiten innerhalb eines Haushalts unabhängig ist. Dadurch würde die Varianzschätzung auf Basis von (4.19) ungeeignet sein. Im Falle einer positiven Korrelation innerhalb eines Haushalts, d.h. zwei Personen eines Haushalts sind eher geneigt gemeinsam umzuziehen als zwei Personen aus unterschiedlichen Haushalten, führt die Annahme der Unabhängigkeit zu einer Unterschätzung des zweiten Terms in (4.19). Gleichzeitig wird auch der dritte Term unterschätzt, der aber, da er subtrahiert wird, für einen Ausgleich in die entgegengesetzte Richtung sorgt.

Um die Varianz bei Abhängigkeiten herzuleiten, werden folgende Größen definiert. U_1 sei Grundgesamtheit von Clustern und U_i die Menge der Einheiten eines Clusters $i \in U_1$. N_1 ist die Anzahl der Cluster. Die Grundgesamtheit der Einheiten ist $U = \{k \in U_i : i \in U_1\}$. Beim Ziehungsverfahren werden alle Einheiten eines Clusters ausgewählt. Wenn $s \subset U$ die Stichprobe der Einheiten ist, dann ist $s_1 = \{i \in U_i : s \cap U_i \neq \{\}\}$ die Stichprobe der Cluster. Das Design-Gewicht für Cluster $i \in U_1$ sei w_i . n_1 sei die erwartete Anzahl der Cluster, bei denen das Untersuchungsmerkmal y für mindestens eine Einheit beobachtet wird. Zusätzlich wird gefordert, dass die Cluster-Größen N_i gleichmäßig beschränkt sind. Damit ist $n/n_1 = O(1)$ und $N/N_1 = O(1)$. Die Annahmen [A2], [A3] und [R1] gelten dementsprechend wenn man n und N durch n_1 bzw. N_1 ersetzt.

Es wird davon ausgegangen, dass für zwei Personen $k \in U_i$ und $l \in U_i$ das Mobilitätsverhalten abhängig ist, d.h. $\theta_{kl} \neq \theta_k \theta_l$. Als Corr_{kl} wird die Korrelation zwischen $1_{\{k \in r\}}$ und $1_{\{l \in r\}}$ definiert. Weiterhin sei $\text{Corr}_{kk} = 1$ und $\text{Corr}_{kl} = 0$ für $k \in U_i$ und $l \in U_j$, $i \neq j$. Dadurch ist die Wahrscheinlichkeit, dass das Untersuchungsmerkmal y sowohl für $k \in s$ als auch $l \in s$ beobachtet wird gegeben durch

$$\theta_{kl} = \theta_k \theta_l \left\{ 1 + \text{Corr}_{kl} \left(\frac{1}{\theta_k} - 1 \right)^{1/2} \left(\frac{1}{\theta_l} - 1 \right)^{1/2} \right\}$$

Nun muss eine Annahme über die Korrelationsstruktur getroffen werden. Hierzu existieren verschiedene Möglichkeiten wie die Korrelationsstruktur parametrisiert werden kann. Diese basieren auf dem GEE-Ansatz (Generalized Estimating Equations) von Liang und Zeger (1986). Der GEE-Ansatz erlaubt die Erweiterung der Generalisierten Linearen Modelle auf Cluster-Daten, d.h. es werden die marginalen Erwartungswerte der korrelierten Zielgrößen wie bei Generalisierten Linearen Regressionsmodellen in Abhängigkeit von Hilfsmerkmalen modelliert. Die Abhängigkeit zwischen den Einheiten wird dann durch die Korrelationsstruktur berücksichtigt. Der einfachste Fall einer Korrelationsstruktur ist $\text{Corr}_{kl} = \alpha_0$ für $k \neq l$ wobei α_0 im Intervall zwischen -1 und 1 liegt, d.h. alle Einheiten haben die gleiche Korrelation.

Die Abhängigkeit zwischen den Einheiten wird für die Schätzung von λ_0 berücksichtigt, indem die Beiträge der Einheiten in (4.3) durch die Korrelation umgewichtet wird. Die Gewichte werden bestimmt, indem die Korrelationsmatrix des Clusters $i \in U_1$ invertiert wird. D.h. es werden Gewichte

w_{Gkl} in Abhängigkeit von der Korrelationsstruktur - festgelegt durch den Parameter α - durch die Lösung des Gleichungssystems

$$\begin{aligned}\sum_{l \in U_i} w_{Gkl}(\alpha) \text{Corr}_{lk} &= 1 \text{ für } k \in U_i \\ \sum_{l \in U_i} w_{Gkl}(\alpha) \text{Corr}_{lk'} &= 0 \text{ für } k, k' \in U_i, k \neq k'\end{aligned}$$

bestimmt. Für diese einfachste Korrelationsstruktur sind die Gewichte

$$w_{Gkl}(\alpha) = \frac{-\alpha}{(1-\alpha)(1+(N_i-1)\alpha)} \quad (4.25)$$

für $k, l \in U_i$ mit $k \neq l$ und

$$w_{Gkk}(\alpha) = \frac{1+(N_i-2)\alpha}{(1-\alpha)(1+(N_i-1)\alpha)} \quad (4.26)$$

für $k \in U_i$.

Unter der Annahme, dass der Parameter $\lambda \in A$ bekannt ist, lässt sich α_0 mit Hilfe der Pearson-Residuen schätzen:

$$\hat{\alpha}(\lambda) = 2 \sum_{U_I} 1_{\{i \in s_I\}} \sum \sum_{U_i, k < l} R_k(\lambda) R_l(\lambda) / (N_i(N_i - 1)) \quad (4.27)$$

wobei die Pearson-Residuen in Abhängigkeit von $\lambda \in A$ wie folgt definiert sind:

$$R_k(\lambda) = \frac{1_{\{k \in r\}} - f_N(\lambda, \mathbf{x}_k)}{\sqrt{f_{N1}(\lambda, \mathbf{x}_k)}} \quad (4.28)$$

Umgekehrt lässt sich der Parameter λ_0 schätzen, indem für bekanntes α eine Nullstelle der Score-Funktion

$$\mathbf{u}(\lambda, \alpha) = \sum_{U_I} 1_{\{i \in s_I\}} \sum \sum_{U_i} w_{Gkl}(\alpha) \mathbf{h}_1(\lambda, \mathbf{x}_k) R_l(\lambda) \quad (4.29)$$

wobei $\mathbf{h}_1(\lambda, \mathbf{x}_k) = \mathbf{x}_k \sqrt{f_{N1}(\lambda, \mathbf{x}_k)}$ ist, bestimmt wird.

Setzt man (4.27) in (4.29) ein, dann ist beim GEE-Ansatz der Schätzer $\hat{\lambda}: \Omega \rightarrow A$ definiert als die Lösung von

$$\mathbf{u}(\lambda, \hat{\alpha}(\lambda)) = 0$$

Mit (4.27) und (4.28) ist $\hat{\alpha}(\hat{\lambda})$ ein Schätzer für die Korrelation, die für die zweite Komponente in (4.19) benutzt werden kann. Im Anhang A.3 wird gezeigt, dass $\sqrt{n_I}(\hat{\lambda} - \lambda_0)$ im p -ten Mittel beschränkt ist, falls $|\hat{\alpha}(\lambda)| \leq a$ für $0 \leq a < 1$ fast sicher gilt. Damit lässt sich wie in 4.3 zeigen, dass $n_I^p \{(\hat{\lambda} - \lambda_0) - \mathbf{F}_M^{-1} \mathbf{u}_M\}^p$ mit $\mathbf{F}_M = \mathbf{F}(\lambda_0, \alpha_0)$ und $\mathbf{u}_M = \mathbf{u}(\lambda_0, \alpha_0)$ in Erwartung beschränkt ist. Hierbei ist

$$\hat{\mathbf{F}}(\lambda, \alpha) = \sum_{U_I} 1_{\{i \in s_I\}} \sum \sum_{U_i} w_{Gkl}(\alpha) \mathbf{h}_1(\lambda, \mathbf{x}_k) \mathbf{h}_1(\lambda, \mathbf{x}_l)'$$

die Fisher Information.

Wie beim PSA-Schätzer unter Unabhängigkeit gilt schließlich

$$\hat{Y}_{\hat{M}} - Y = (\hat{Y}_M - Y) + \mathbf{D}'_M \hat{\mathbf{F}}_M^{-1} \mathbf{u}_M + O_p(n^{-1})$$

mit

$$\begin{aligned} \mathbf{D}_M &= -N^{-1} \sum_{U_1} \sum_{U_i} y_k \mathbf{x}_k (1 - \theta_k) \\ \hat{\mathbf{F}}_M &= \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i} w_{Gkl}(\alpha_0) \mathbf{h}_1(\lambda_0, \mathbf{x}_k) \mathbf{h}_1(\lambda_0, \mathbf{x}_l)' \end{aligned}$$

wobei $f_L = \text{id}$ angenommen wird. Unter Beachtung von $E[\hat{Y}_M \mathbf{u}_M] = -\mathbf{D}_M$ gilt für die Varianz

$$V[\hat{Y}_{\hat{M}} - Y] = V[\hat{Y}_M - Y] - E[\mathbf{D}'_M \hat{\mathbf{F}}_M^{-1} \mathbf{D}_M] + O(n^{-2})$$

Mit (2.8) ist die Varianz

$$\begin{aligned} V[\hat{Y}_{\hat{M}}] &= \frac{1}{N^2} \sum \sum_{U_1} \Delta_{ij} \left(\sum_{U_i} w_{iy} y_k \right) \left(\sum_{U_j} w_{lj} y_l \right) + \\ &\quad \frac{1}{N^2} E \left[\sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i} \Delta_{Mkl} (w_{Mk} y_k) (w_{Ml} y_l) \right] - E[\mathbf{D}'_M \hat{\mathbf{F}}_M^{-1} \mathbf{D}_M] + O(n^{-2}) \end{aligned} \quad (4.30)$$

mit $w_{Mk} = w_{iy} \{1 + \exp(-\lambda_0' \mathbf{x}_k)\}$, $k \in U_i$, $\theta_k = \{1 + \exp(-\lambda_0' \mathbf{x}_k)\}^{-1}$ und

$$\Delta_{Mkl} = \alpha_0 \theta_k \theta_l \sqrt{\left(\frac{1}{\theta_k} - 1\right) \left(\frac{1}{\theta_l} - 1\right)}$$

für $k, l \in U_i$ mit $k \neq l$. Der Varianzschätzer ist

$$\begin{aligned} \hat{V}_M &= \frac{1}{N^2} \sum \sum_{U_1} \frac{\Delta_{ij}}{\pi_{ij}} \left(\sum_{U_i} 1_{\{k \in r\}} \hat{w}_{Mk} y_k \right) \left(\sum_{U_j} 1_{\{l \in r\}} \hat{w}_{Ml} y_l \right) + \\ &\quad \frac{1}{N^2} \sum_{U_1} 1_{\{i \in s_1\}} \pi_{ii} \sum \sum_{U_i} 1_{\{k, l \in r\}} \check{\Delta}_{Mkl} (\hat{w}_{Mk} y_k) (\hat{w}_{Ml} y_l) - \hat{\mathbf{D}}'_M \hat{\mathbf{F}}_M^{-1} \hat{\mathbf{D}}_M \end{aligned} \quad (4.31)$$

wobei $\hat{w}_{Mk} = w_{iy} \{1 + \exp(-\hat{\lambda}' \mathbf{x}_k)\}$, $\hat{\theta}_k = \{1 + \exp(-\hat{\lambda}' \mathbf{x}_k)\}^{-1}$,

$$\begin{aligned} \hat{\mathbf{D}}_{\hat{M}} &= -N^{-1} \sum_{U_1} \sum_{U_i} 1_{\{k \in r\}} w_{iy} y_k \mathbf{x}_k (1 - \hat{\theta}_k) \\ \hat{\mathbf{F}}_{\hat{M}} &= \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i} w_{Gkl}(\hat{\alpha}) \mathbf{h}_1(\hat{\lambda}, \mathbf{x}_k) \mathbf{h}_1(\hat{\lambda}, \mathbf{x}_l)' \end{aligned}$$

und

$$\check{\Delta}_{Mkl} = \hat{\alpha} \sqrt{\left(\frac{1}{\hat{\theta}_k} - 1\right) \left(\frac{1}{\hat{\theta}_l} - 1\right)} / \left\{ 1 + \hat{\alpha} \sqrt{\left(\frac{1}{\hat{\theta}_k} - 1\right) \left(\frac{1}{\hat{\theta}_l} - 1\right)} \right\}$$

für $k, l \in U_i$ mit $k \neq l$.

5 Simulation

Im folgenden werden die unterschiedlichen Verfahren zur Korrektur des Nonresponse Verzerrung anhand des SOEP miteinander verglichen. Die Beurteilung der verschiedenen Verfahren erfolgt mit Hilfe verschiedener Qualitätskriterien. Hierzu wird der relative Bias des Kalibrationsschätzers und des PSA-Schätzers sowie der relative Bias des geschätzten Standardfehlers betrachtet.

Die Berechnung des relativen Bias basiert auf 1'000 aus dem SOEP gezogenen Stichproben. Jede Stichprobe wurde dabei so gezogen, dass das Ziehungsverfahren des Mikrozensus abgebildet wurde. Dazu wurde das SOEP in Schichten eingeteilt, wobei die Schichten den Bundesländern entsprachen. Um der Klumpung im Mikrozensus Rechnung zu tragen, wurden als Klumpen PLZ-Gebiete definiert.²⁷ Danach wurden innerhalb jeder Schicht 20% der PLZ-Gebiete ausgewählt. Anschließend wurde die Mobilität durch ein nicht-parametrisches Verfahren simuliert. 500 Stichproben wurden dabei mit Zurücklegen aus jeder der 1'000 Stichproben gezogen. Hierbei ist zu beachten, dass die Verwendung des Bootstraps im Design-basierten Ansatz mit methodischen Schwierigkeiten verbunden ist. Dies liegt daran, dass die Verteilung der Mobilität nicht korrekt abgebildet wird. Allerdings kann gezeigt werden, dass durch Verwendung einer Korrektur die Bootstrap-Varianz asymptotisch mit der theoretischen Varianz übereinstimmt. Die Korrektur ist unabhängig von dem zugrunde gelegtem Nonresponse-Modell. Dies wird in Abschnitt 5.3 gezeigt.

Bevor auf das Korrekturverfahren näher eingegangen wird, werden der Mikrozensus und das SOEP kurz beschrieben. Die Ergebnisse werden in Abschnitt 5.4 dargestellt.

5.1 Das Erhebungsdesign des Mikrozensus

Nach Angaben des Statistischen Bundesamtes werden die zu befragenden Haushalte und Personen im MZ im Rahmen einer Flächenstichprobe ermittelt. Dazu werden aus dem Bundesgebiet Flächen (Auswahlbezirke) ausgewählt, in denen alle Haushalte und Personen befragt werden. Die Auswahlbezirke bestehen durchschnittlich aus 6 bis 12 benachbarten Wohnungen, die in eine Gebäudegruppe oder innerhalb eines grösseren Gebäudes liegen, siehe Herther-Eschweiler (2003).

Beim Mikrozensus handelt sich um eine einstufig geschichtete Klumpenstichprobe, wobei Auswahlbezirke die Klumpen sind. Die Schichtung erfolgt sowohl regional als auch fachlich. Die regionale Schichten sind im Allgemeinen Zusammenlegungen von Kreisen und kreisfreien Städten. Ihre durchschnittliche Einwohnerzahl beziffert sich auf durchschnittlich 250'000 Einwohner, wobei Großstädte mit 200'000 oder mehr Einwohnern eine eigene regionale Schicht bilden und sehr große Städte wie beispielsweise Berlin und Hamburg in mehrere regionale Schichten unterteilt sind. Die regionalen Schichten sind wiederum in fünf Teilschichten unterteilt, den sogenannten vier Gebäudegrößenklassen und der Neubauschicht (fachliche Schichtung). Die erste Gebäudegrößenklassen enthält Gebäude mit 1 bis 4 Wohnungen, die zweite Gebäude mit 5 bis 10 Wohnungen, die dritte Gebäude mit 11 oder mehr Wohnungen, die vierte enthält Gebäude mit vermuteten Gemeinschaftsunterkünften und die fünfte die Neubauten (Neubauschicht).

Innerhalb dieser $214 \times 5 = 1070$ Schichten sind die Auswahlbezirke angeordnet. Je 100 Auswahlbezirke bilden für das Ziehungsverfahren eine Schicht. Vier in der Reihenfolge der Sortierung aufeinander folgende Schichten bilden einen Block. Den Schichten eines Blocks werden die Zahlen 1 bis 4 zufällig

²⁷Die PLZ-Gebiete wurde von Jan Goebel (DIW) zur Verfügung gestellt.

zugeordnet. Auswahlbezirke mit gleicher Zonenummer gehören dem gleichen Rotationsviertel an, womit eine Zerlegung der 1% - Stichproben in vier Rotationsviertel zu je 0,25% gegeben ist.

Nach Angaben des Mikrozensusgesetzes werden die in den Auswahlbezirken wohnenden Haushalte und Personen in vier aufeinander folgenden Jahren befragt. In jedem Jahr scheidet ein Viertel der Stichprobe aus der Erhebung aus und ein neues Viertel wird erstmals einbezogen (sogenanntes Rotationsverfahren). Der MZ ist somit als Wiederholungsbefragung mit einer teilweisen Überlappung der Erhebungseinheiten angelegt. Die Stichprobe eines Jahres setzt sich danach aus Auswahlbezirken zusammen, die jeweils zu einem Viertel entweder nicht mehr oder ein, zwei oder drei weitere Male in die Erhebung einbezogen werden.

5.2 Das Sozio-oekonomische Panel

Das Sozio-oekonomische Panel (SOEP) ist eine seit 1984 jährlich durchgeführte Befragung von Privathaushalten in Deutschland. Die erhobenen Daten geben unter anderem Auskunft über Einkommen, Erwerbstätigkeit, Bildung und Gesundheit. Da jedes Jahr weitgehend dieselben Personen sowie Haushalte zu den gleichen Themen befragt werden, können nicht nur langfristige gesellschaftliche Trends, sondern auch die gruppenspezifische Entwicklung von Lebensläufen besonders gut analysiert werden.

1984 wurde die erste Erhebung des SOEP mit rund 6'000 Haushalten durchgeführt. Die Zielpopulation der Starterhebung war durch die westdeutschen Haushalte bzw. Personen, die in diesen Haushalten lebten, definiert. Diese Starterhebung des SOEP bestand aus zwei getrennten Stichproben: Teilstichprobe A (4'500 Haushalte) umfasst Haushalte mit einem deutschen Haushaltsvorstand während die Teilstichprobe B (1'500 Haushalte) die Haushalte mit einem ausländischen (italienischen, spanischen, türkischen, griechischen oder ex-jugoslawischen) Haushaltsvorstand umfasst. Im Juni 1990 wurde das SOEP auf das Gebiet der ehemaligen DDR ausgeweitet (Stichprobe C mit 2'000 Haushalten). Um den Aspekt der Zuwanderung nach Deutschland zu berücksichtigen, wurde 1994/1995 eine spezielle Zuwanderer Stichprobe erhoben (Stichprobe D mit etwa 500 Haushalten). Diese umfasst Personen (überwiegend Aus- und Übersiedler), die in den Jahren von 1984 bis einschließlich 1994/1995 aus dem Ausland zugewandert sind. In den Folgejahren wurden weitere Ergänzungsstichproben gezogen, die vor allem der Stabilisierung der Fallzahlen dienen.

Trotz vieler gemeinsamer Design-Elemente unterscheidet sich die Auswahl der Starthaushalte in den vier Teilstichproben A, B, C und D. Die Weiterverfolgungsregeln der Mitglieder der Starthaushalte sind jedoch in allen Teilstichproben identisch. Weiterhin ist allen vier Teilstichproben die Zweistufigkeit des Auswahlverfahrens gemeinsam. Zusätzlich wurde sowohl für die Auswahl der Primäreinheiten als auch der Sekundäreinheiten (Haushalte) bei allen drei Stichproben größenproportionales systematisch Ziehen mit Intervall und zufälliger Startzahl gewählt, siehe Spiess (2005).

Für die Simulation werden die Jahre 1996 bis 1999 der Stichproben A bis D verwendet.

Trotz der methodischen Unterschiede zwischen des Mikrozensus und des SOEP wurde in Bašić (2008) gezeigt, dass das Ausmaß der räumlichen Mobilität zwischen dem SOEP und dem Mikrozensus vergleichbar ist.

5.3 Theoretische Herleitung der Simulation

Das SOEP enthält durch das Konzept der Weiterverfolgung für alle räumlich immobilen und mobilen Personen Ausprägungen von Untersuchungsmerkmalen, die im Mikrozensus nur für räumlich

immobile Personen beobachtbar sind. Dadurch lässt sich der Mittelwert oder der Gesamtwert eines Untersuchungsmerkmal y anhand des SOEPs berechnen und dabei ist dieser nicht durch die Mobilität beeinträchtigt. Damit ist das SOEP für eine Simulationsstudie geeignet, wobei die Personen des SOEPs als Grundgesamtheit U verwendet werden. Im Folgenden wird gezeigt, wie durch ein nicht-parametrisches Verfahren die Mobilität simuliert werden kann. Das unten gezeigte nicht-parametrische Verfahren hat gegenüber einem parametrischen Verfahren den Vorteil, dass mögliche Einflüsse von Merkmalen \mathbf{x} und y nicht außer acht gelassen werden, da ein Modell immer eine Vereinfachung der Realität darstellt und möglicherweise das eine oder andere Korrekturverfahren bevorzugen könnte.

Zunächst erfolgt die Beschreibung des Simulationsverfahrens. Hierzu sei \mathcal{S} die Menge der möglichen Stichproben aus U . Das Stichprobendesign sei wie beim Mikrozensus eine geschichtete Klumpenstichprobe, wobei die Schichten Bundesländer und die Klumpen Postleitzahlengebiete sind. Neben der Struktur für das Stichprobendesign enthält das SOEP eine Realisierung $r \in \mathcal{S}$ derjenigen Personen, die an ihrem Wohnort verblieben sind. Allgemein kann das Ziehungsverfahren $p_D(\cdot|r) : \mathcal{S} \rightarrow [0, 1]$ bei der Simulationsstudie informativ bezüglich der Mobilität sein. Die Mobilität im SOEP ist selbst eine Realisierung durch einen unbekanntem Prozeß $q : \mathcal{S} \rightarrow [0, 1]$. Somit gilt

$$p_D(s) = \sum_{\mathcal{S}} p_D(s|r)q(r)$$

Hierbei wird davon ausgegangen, dass die im SOEP beobachtbare Mobilität ein Abbild der Mobilität in der deutschen Bevölkerung darstellt. Für $r \subset s$ und $p_D(s) > 0$ ist der auf der Stichprobe bedingte Prozeß

$$q(r|s) = \frac{1}{p_D(s)} \sum_{\{r_U \supset r\}} p_D(s|r_U)q(r_U)$$

und $q(r|s) = 0$ sonst. Dies bedeutet, dass nach Ziehung einer Stichprobe aus dem SOEP, eine Realisierung der räumlich immobilen Personen $r \in \mathcal{S}$ gegeben die Stichprobe $s \in \mathcal{S}$ beobachtet wird. Da mit dem SOEP nur eine Realisierung beobachtbar ist, lässt sich mit Hilfe des Bootstraps, Efron (1979), die Mobilität simulieren.

Damit eine mögliche Abhängigkeit der räumlichen Mobilität einzelner Personen innerhalb eines Haushalts nicht zerstört wird, ist es sinnvoll beim Bootstrap Haushalte zu ziehen. Sei N_I die Anzahl der Haushalte im SOEP, U_I die Grundgesamtheit der Haushalte und U_i die Menge der Personen des Haushaltes $i \in U_I$. Durch das Ziehungsverfahren werden alle Einheiten eines Haushaltes ausgewählt. Die Stichprobe der Haushalte ist $s_I = \{i \in U_I : s \cap U_i \neq \emptyset\}$. Die Auswahlwahrscheinlichkeiten seien π_{i_i} und π_{i_j} für $i, j \in U_I$. Für alle Personen des Haushaltes $i \in U_I$ ist das Design-Gewicht $w_k = \pi_{i_i}^{-1}$, $k \in U_i$. Der Bootstrap simuliert die multinomiale Verteilung. Sei $\mathbb{M} = \{1, 2, \dots, N_I\}$ die Menge der Haushaltsindizes. Falls die Stichprobe $s \in \mathcal{S}$ gezogen wurde, steht auch die Anzahl der zu ziehenden Haushalte $n_I(s)$ fest. Beim Bootstrap der Haushalte ist die multinomiale Verteilung $p_M(\cdot|s, r) : \mathbb{M}^{N_I} \rightarrow [0, 1]$, durch

$$p_M(m|s, r) = \frac{n_I(s)}{m_1(s)!m_2(s)! \cdots m_{N_I}(s)!} \left(\frac{1}{n_I(s)} \right)^{n_I(s)}$$

gegeben, falls $\sum_{U_I} m_i(s) = n_I(s)$. Ansonsten gilt $p_M(m|s, r) = 0$. Zusätzlich gilt für Haushalte, die nicht durch das Ziehungsverfahren ausgewählt wurden, $m_i(s) = 0$.

Die Simulation basiert auf der im SOEP beobachtbaren Mobilität, dem Ziehungsverfahren wie beim Mikrozensus und des Bootstraps der Haushalte. Die Ergebnismenge ist $\Omega = \mathcal{S} \times \mathcal{S} \times \mathbb{M}^{N_1}$. Das Wahrscheinlichkeitsmaß P ist mit den obigen Definitionen für $(s, r, m) \in \Omega$ durch

$$p(s, r, m) = p_M(m|s, r)p_D(s|r)q(r)$$

gegeben. $\mathcal{H} = \{A \times \mathcal{S} \times \mathbb{M}^{N_1} : A \subset \mathcal{S}\}$ ist die verfügbare Information nach Auswahl der Stichprobe. Das SOEP enthält eine Realisierung der Mobilität. Dennoch wird die Mobilität als Zufallsgröße betrachtet (zum Zeitpunkt der Erhebung des SOEPs ist sie noch nicht bekannt). Dann wird eine Stichprobe wie im Mikrozensus gezogen. Anschließend werden Bootstrap-Stichproben mit Zurücklegen gezogen. Damit ist

$$\mathbb{E}[1_{\{k \in r \cap s\}}|\mathcal{H}] = \theta_k 1_{\{k \in s\}} \text{ und } \mathbb{E}[1_{\{k, l \in r \cap s\}}|\mathcal{H}] = \theta_{kl} 1_{\{k, l \in s\}}$$

für $k, l \in U$. $\mathcal{F} = \{A \times \mathbb{M}^{N_1} : A \subset \mathcal{S} \times \mathcal{S}\}$ ist die verfügbare Information auf Basis der Mobilität und der Stichprobenziehung. $M_i : \mathbb{M}^{N_1} \rightarrow \mathbb{M}$ mit $M_i(m) = m_i$ gibt an, wie oft Haushalt $i \in U_1$ beim Bootstrap ausgewählt wird. Es gilt

$$\mathbb{E}[M_i|\mathcal{F}] = 1_{\{i \in s_1\}}, \text{ V}[M_i|\mathcal{F}] = 1_{\{i \in s_1\}}(1 - 1/n_1) \text{ und } \text{Cov}[M_i, M_j|\mathcal{F}] = -1_{\{i \in s_1\}}/n_1$$

fast sicher.

Für theoretische Überlegungen seien zunächst $q : \mathcal{S} \rightarrow [0, 1]$ und die Wahrscheinlichkeiten θ_k und θ_{kl} für $k, l \in U$ bekannt. Mit $w_{Mk} = (\pi_k \theta_k)^{-1}$ ist der Bootstrap-Schätzer für den Mittelwert in der Grundgesamtheit auf Basis (2.7)

$$\hat{Y}_M^* = \frac{1}{N} \sum_{U_1} M_i \sum_{U_i} 1_{\{k \in r \cap s\}} w_{Mk} y_k \tag{5.1}$$

Dieser hat den Erwartungswert

$$\mathbb{E}[\hat{Y}_M^*|\mathcal{F}] = \hat{Y}_M = \frac{1}{N} \sum_{U_1} \sum_{U_i} 1_{\{k \in r \cap s\}} w_{Mk} y_k$$

und ist somit ein erwartungstreuer Schätzer für den Mittelwert in der Grundgesamtheit. Die Varianz, die durch den Bootstrap erzeugt wird, ist

$$\begin{aligned} \text{V}[\hat{Y}_M^*|\mathcal{F}] &= \frac{1}{N^2} \sum_{U_1} \sum \sum_{U_i} (1 - 1/n_1) 1_{\{k, l \in r \cap s\}} (w_{Mk} y_k)(w_{Ml} y_l) \\ &\quad - \frac{1}{N^2} \sum \sum_{U_1, i \neq j} \sum_{U_i} \sum_{U_j} 1/n_1 1_{\{k, l \in r \cap s\}} (w_{Mk} y_k)(w_{Ml} y_l) \end{aligned}$$

und hat den Erwartungswert

$$\begin{aligned}
\mathbb{E}[V[\hat{Y}_M^*|\mathcal{F}]|\mathcal{H}] &= \frac{1}{N^2} \sum_{U_I} 1_{\{i \in s_I\}} \sum \sum_{U_i} (1 - 1/n_I) \theta_{kl} (w_{Mk} y_k) (w_{MI} y_l) \\
&\quad - \frac{1}{N^2} \sum \sum_{U_I, i \neq j} 1_{\{i, j \in s_I\}} \sum_{U_i} \sum_{U_j} 1/n_I (w_k y_k) (w_l y_l) \\
&= \frac{1}{N^2} \sum_{U_I} 1_{\{i \in s_I\}} \sum \sum_{U_i} (1 - 1/n_I) (\theta_{kl} - \theta_k \theta_l) (w_{Mk} y_k) (w_{MI} y_l) \\
&\quad + \frac{1}{N^2} \sum_{U_I} 1_{\{i \in s_I\}} \left(\sum_{U_i} w_k y_k - n_I^{-1} N \hat{Y}_{HT} \right)^2
\end{aligned}$$

Die Varianz stimmt nicht in Erwartung mit der Varianz der Verteilungsfunktion $P(\{\hat{Y}_M - \hat{Y}_{HT} \in C\}|\mathcal{H})$, $C \in \mathcal{B}(\mathbb{R})$, überein. Der letzte Term ist die Bootstrap Varianz des Horvitz-Thompson Schätzers. Betrachtet man die Differenz $\hat{Y}_M^* - \hat{Y}_{HT}^*$, wobei

$$\hat{Y}_{HT}^* = \frac{1}{N} \sum_{U_I} M_i \sum_{U_i} 1_{\{k \in s\}} w_k y_k \tag{5.2}$$

der Bootstrap-Schätzer des Horvitz-Thompson Schätzers ist. Dann ist

$$\mathbb{E}[\hat{Y}_M^* - \hat{Y}_{HT}^*|\mathcal{F}] = \hat{Y}_M - \hat{Y}_{HT}$$

und die Varianz

$$\hat{V}_{NR}^* = V[\hat{Y}_M^* - \hat{Y}_{HT}^*|\mathcal{F}] = V[\hat{Y}_M^*|\mathcal{F}] + V[\hat{Y}_{HT}^*|\mathcal{F}] - 2\text{Cov}[\hat{Y}_M^*, \hat{Y}_{HT}^*|\mathcal{F}]$$

hat den Erwartungswert

$$\mathbb{E}[\hat{V}_{NR}^*|\mathcal{H}] = \frac{1}{N^2} \sum_{U_I} 1_{\{i \in s_I\}} \sum \sum_{U_i} (1 - 1/n_I) (\theta_{kl} - \theta_k \theta_l) (w_{Mk} y_k) (w_{MI} y_l)$$

Die Varianz $V[\hat{Y}_M^* - \hat{Y}_{HT}^*|\mathcal{F}]$ stimmt bis auf den Faktor $(n_I - 1)/n_I$ in Erwartung mit der theoretischen durch die Mobilität erzeugten Varianz überein. Die Idee ist daher, als Approximation für die Verteilungsfunktion $P(\{\hat{Y}_M - \hat{Y}_{HT} \in C\}|\mathcal{H})$, $C \in \mathcal{B}(\mathbb{R})$, die Verteilungsfunktion $P(\{(\hat{Y}_M^* - \hat{Y}_{HT}^*) - (\hat{Y}_M - \hat{Y}_{HT}) \in C\}|\mathcal{F})$, $C \in \mathcal{B}(\mathbb{R})$, zu benutzen. Beide Größen sind Zufallsvariablen.

Der Unterschied zwischen $\mathbb{E}[\hat{V}_{NR}^*|\mathcal{H}]$ und $V[\hat{Y}_M|\mathcal{H}]$ ist

$$\mathbb{E}[\hat{V}_{NR}^*|\mathcal{H}] - V[\hat{Y}_M|\mathcal{H}] = -\frac{1}{n_I} V[\hat{Y}_M|\mathcal{H}]$$

bzw.

$$n_I \left(\mathbb{E}[\hat{V}_{NR}^*|\mathcal{H}] - V[\hat{Y}_M|\mathcal{H}] \right) = -\frac{1}{n_I} \mathbb{E}[n_I (\hat{Y}_M - \hat{Y}_{HT})^2|\mathcal{H}]$$

Die Anzahl ausgewählter Haushalte n_I ist eine Zufallsgröße. Es wird aber davon ausgegangen, dass das Ziehungsverfahren so angelegt ist, dass eine Mindestanzahl an Haushalten fast sicher ausgewählt wird. Zusätzlich soll die Mindestanzahl gegen Unendlich wachsen. Somit kann man davon ausgehen, dass n_I^{-1} fast sicher gegen Null konvergiert. Außerdem sollte jeder Haushalt aus mindestens einer Person bestehen und die Haushaltsgröße gleichmäßig beschränkt sein. Damit konvergiert auch die erwartete Anzahl der Personen n fast sicher gegen Unendlich. Mit den Annahmen [A2] und [A3] konvergiert daher

der Erwartungswert der rechten Seite gegen Null und somit auch die linke Seite. Folglich konvergiert die (mit n_1 stabilisierte) Bootstrap-Varianz in Erwartung gegen die (mit n_1 stabilisierte) theoretische Varianz. Weiterhin gilt mit den Annahmen [A2] und [A3]

$$\mathbb{E}[n_1(\hat{Y}_M^* - \hat{Y}_{HT}^*)^2 | \mathcal{H}] = \left(2 - \frac{1}{n_1}\right) \mathbb{E}[n_1(\hat{Y}_M - \hat{Y}_{HT})^2 | \mathcal{H}]$$

und daher $\sqrt{n_1}(\hat{Y}_M^* - \hat{Y}_{HT}^*) = O_p(1)$.

$\hat{V}_{NR}^* = V[\hat{Y}_M^* - \hat{Y}_{HT}^* | \mathcal{F}]$ ist ein asymptotisch erwartungstreuer Schätzer für die Varianz $V[\hat{Y}_M | \mathcal{H}]$, falls der Schätzer \hat{Y}_M^* auf Basis der Wahrscheinlichkeiten θ_k und θ_{kl} , $k, l \in U$, gebildet werden könnte. Der berechenbare Schätzer ist aber (2.10) mit gegebener funktionaler Form $f_M: A \times B \rightarrow \mathbb{R}$ und Parameter $\lambda_0: \mathcal{S} \rightarrow A$.²⁸ Für jede Bootstrap-Stichprobe wird λ_0 durch $\hat{\lambda}^*: \Omega \rightarrow A$ geschätzt, $\hat{w}_{Mk}^* = w_k f_M(\hat{\lambda}^*, \mathbf{x}_k)$ berechnet und

$$\hat{Y}_{M^*}^* = \frac{1}{N} \sum_{U_1} M_i \sum_{U_i} 1_{\{k \in r \cap s\}} \hat{w}_{Mk}^* y_k \quad (5.3)$$

bestimmt. Der Bootstrap ist unter gegebener funktionaler Form f_M asymptotisch korrekt, falls die Verteilungsfunktion $P(\{\hat{Y}_M - Z \in C\} | \mathcal{H})$, $C \in \mathcal{B}(\mathbb{R})$ durch die Verteilungsfunktion

$$P(\{(\hat{Y}_{M^*}^* - Z^*) - (\hat{Y}_M - Z) \in C\} | \mathcal{F}), \quad C \in \mathcal{B}(\mathbb{R})$$

approximiert werden kann. Hierbei ist $Z^*: \Omega \rightarrow A$ eine vom Verfahren abhängige Zufallsgröße und $Z = \mathbb{E}[Z^* | \mathcal{F}]$.

Seien f_{M1} und f_{M2} die ersten und zweiten Ableitungen von f_M . Außerdem seien $f_{K1}(\lambda, \mathbf{x}_k) = f_{M1}(\lambda, \mathbf{x}_k)/f_M(\lambda_0, \mathbf{x}_k)$, $f_{K2}(\lambda, \mathbf{x}_k) = f_{M2}(\lambda, \mathbf{x}_k)/f_M(\lambda_0, \mathbf{x}_k)$ und das Wegintegral

$$F_{K2}(\lambda_P(\lambda, \cdot), \cdot) = \int_0^1 f_{K2}(\lambda_P(\lambda, a), \cdot) da$$

mit $\lambda_P(\lambda, a) = \lambda_0 + a(\lambda - \lambda_0)$ gegeben. Mit

$$\begin{aligned} \hat{\mathbf{D}}_M^* &= N^{-1} \sum_{U_1} \sum_{U_i} M_i 1_{\{k \in r \cap s\}} w_{Mk} y_k f_{K1}(\lambda_0, \mathbf{x}_k) \text{ und} \\ \hat{\mathbf{R}}_M^*(\lambda) &= N^{-1} \sum_{U_1} \sum_{U_i} M_i 1_{\{k \in r \cap s\}} w_{Mk} y_k F_{K2}(\lambda_P(\lambda, \cdot), \mathbf{x}_k) \end{aligned}$$

lässt sich (5.3) darstellen als

$$\hat{Y}_{M^*}^* = \hat{Y}_M^* + \hat{\mathbf{D}}_M^{*'}(\hat{\lambda}^* - \lambda_0) + \frac{1}{2}(\hat{\lambda}^* - \lambda_0)' \hat{\mathbf{R}}_M(\hat{\lambda}^*)(\hat{\lambda}^* - \lambda_0)$$

Es ist zu beachten, dass bei Anwendung des Kalibrationsschätzers die zweiten Ableitungen Null sind und daher auch der letzte Term Null ist.

Angenommen es gilt $\sqrt{n_1}(\hat{\lambda}^* - \lambda_0) = O_p(1)$. Dies gilt für alle betrachteten Schätzverfahren und wird weiter unten gezeigt. Dann ist

$$\hat{Y}_{M^*}^* = \hat{Y}_M^* + \hat{\mathbf{D}}_M^{*'}(\hat{\lambda}^* - \lambda_0) + O_p(n^{-1}) \quad (5.4)$$

²⁸Man beachte, dass sich der Modell-Parameter von Stichprobe zu Stichprobe unterscheiden darf.

Zunächst wird der Kalibrationssschätzer aus Abschnitt 2 betrachtet. Hierbei ist $\hat{\mathbf{X}}^* : \Omega \rightarrow \mathbb{R}^d$ eines der folgenden Schätzer

$$\text{Fall C1: } \hat{\mathbf{X}}_{\text{HT}}^* = \frac{1}{N} \sum_{U_1} \sum_{U_i} M_i 1_{\{k \in s\}} w_k \mathbf{x}_k$$

$$\text{Fall B1: } \hat{\mathbf{X}}_{\text{B1}}^* = \begin{pmatrix} \mathbf{X}_A \\ \hat{\mathbf{X}}_{\text{BHT}}^* \end{pmatrix}$$

$$\text{Fall B2: } \mathbf{X} = \frac{1}{N} \sum_{U_1} \sum_{U_i} \mathbf{x}_k$$

$$\text{Fall A1: } \hat{\mathbf{X}}_{\text{A1}}^* = \begin{pmatrix} \mathbf{X}_A \\ \hat{\mathbf{X}}_{\text{BCAL}}^* \end{pmatrix}$$

wobei die einzelnen Größen von $\hat{\mathbf{X}}_{\text{B1}}^*$ und $\hat{\mathbf{X}}_{\text{A1}}^*$ durch

$$\hat{\mathbf{X}}_{\text{BCAL}}^* = \hat{\mathbf{X}}_{\text{BHT}}^* + \hat{\mathbf{B}}_A^{\text{B}'} (\mathbf{X}_A - \hat{\mathbf{X}}_{\text{AHT}}^*)$$

$$\hat{\mathbf{B}}_A^{\text{B}} = (\sum_{U_1} \sum_{U_i} 1_{\{k \in s\}} w_k \mathbf{x}_{Ak} \mathbf{x}'_{Ak})^{-1} \sum_{U_1} \sum_{U_i} 1_{\{k \in s\}} w_k \mathbf{x}_{Ak} \mathbf{x}'_{Bk}$$

mit

$$\hat{\mathbf{X}}_{\text{AHT}}^* = N^{-1} \sum_{U_1} \sum_{U_i} 1_{\{k \in s\}} w_k \mathbf{x}_{Ak}$$

$$\hat{\mathbf{X}}_{\text{BHT}}^* = N^{-1} \sum_{U_1} \sum_{U_i} M_i 1_{\{k \in s\}} w_k \mathbf{x}_{Bk}$$

gegeben sind. Unter dem Kalibrationsmodell, d.h. $f_{\text{M}}(\lambda, \mathbf{x}_k) = 1 + \lambda' \mathbf{x}_k$ ist λ_0 implizit durch

$$\lambda_0 = N \left(\sum_{U_1} \sum_{U_i} M_i 1_{\{k \in r \cap s\}} d_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} (\hat{\mathbf{X}}_{\text{M}}^* - \hat{\mathbf{X}}_{\text{NR}}^*)$$

gegeben. Der Schätzer für λ_0 ist

$$\hat{\lambda}^* = N \left(\sum_{U_1} \sum_{U_i} M_i 1_{\{k \in r \cap s\}} w_k \mathbf{x}_k \mathbf{x}'_k \right)^{-1} (\hat{\mathbf{X}}^* - \hat{\mathbf{X}}_{\text{NR}}^*)$$

Die Größen $\hat{\mathbf{X}}_{\text{NR}}^*$ und $\hat{\mathbf{X}}_{\text{M}}^*$ sind hierbei wie folgt definiert.

$$\hat{\mathbf{X}}_{\text{NR}}^* = N^{-1} \sum_{U_1} \sum_{U_i} M_i 1_{\{k \in s \cap r\}} w_k \mathbf{x}_k$$

$$\hat{\mathbf{X}}_{\text{M}}^* = N^{-1} \sum_{U_1} \sum_{U_i} M_i 1_{\{k \in s \cap r\}} w_{\text{M}k} \mathbf{x}_k$$

Damit ist (5.4)

$$\hat{Y}_{\hat{\mathbf{M}}^*}^* = \hat{Y}_{\text{M}}^* + (\hat{\mathbf{X}}^* - \hat{\mathbf{X}}_{\text{M}}^*)' \hat{\mathbf{B}}_{\text{NR}}^* \text{ mit}$$

$$\hat{\mathbf{B}}_{\text{NR}}^* = (\sum_{U_1} \sum_{U_i} M_i 1_{\{k \in s \cap r\}} w_k \mathbf{x}_k \mathbf{x}'_k)^{-1} \sum_{U_1} \sum_{U_i} M_i 1_{\{k \in s \cap r\}} w_k \mathbf{x}_k y_k$$

oder auch

$$\begin{aligned} \hat{Y}_{\hat{\mathbf{M}}^*}^* = & \hat{Y}_{\text{M}}^* - (\hat{\mathbf{X}}_{\text{M}}^* - \hat{\mathbf{X}}_{\text{HT}}^*)' \mathbf{B}_{\text{NR}} + (\hat{\mathbf{X}}^* - \hat{\mathbf{X}}_{\text{HT}}^*)' \hat{\mathbf{B}}_{\text{NR}} + \\ & (\hat{\mathbf{X}}_{\text{HT}}^* - \hat{\mathbf{X}}_{\text{M}}^*)' (\hat{\mathbf{B}}_{\text{NR}}^* - \mathbf{B}_{\text{NR}}) + (\hat{\mathbf{X}}^* - \hat{\mathbf{X}}_{\text{HT}}^*)' (\hat{\mathbf{B}}_{\text{NR}}^* - \hat{\mathbf{B}}_{\text{NR}}) \end{aligned}$$

wobei

$$\begin{aligned}\hat{\mathbf{B}}_{\text{NR}} &= (\sum_{U_1} \sum_{U_i} 1_{\{k \in s \cap r\}} w_k \mathbf{x}_k \mathbf{x}'_k)^{-1} \sum_{U_1} \sum_{U_i} 1_{\{k \in s \cap r\}} w_k \mathbf{x}_k y_k \text{ und} \\ \mathbf{B}_{\text{NR}} &= (\sum_{U_1} \sum_{U_i} (1 + \lambda'_0 \mathbf{x}_k)^{-1} \mathbf{x}_k \mathbf{x}'_k)^{-1} \sum_{U_1} \sum_{U_i} (1 + \lambda'_0 \mathbf{x}_k)^{-1} \mathbf{x}_k y_k\end{aligned}$$

ist.

Mit der Annahme [A2] und [R1] aus Abschnitt 2 ist

$$\hat{Y}_{\hat{M}^*}^* = \hat{Y}_M^* - (\hat{\mathbf{X}}_M^* - \hat{\mathbf{X}}_{\text{HT}}^*)' \mathbf{B}_{\text{NR}} + (\hat{\mathbf{X}}^* - \hat{\mathbf{X}}_{\text{HT}}^*)' \hat{\mathbf{B}}_{\text{NR}} + O_p(n^{-1})$$

Somit gilt

$$\hat{Y}_{\hat{M}^*}^* - \hat{Y}_{\text{HT}}^* - (\hat{\mathbf{X}}^* - \hat{\mathbf{X}}_{\text{HT}}^*)' \hat{\mathbf{B}}_{\text{NR}} = (\hat{Y}_M^* - \hat{Y}_{\text{HT}}^*) - (\hat{\mathbf{X}}_M^* - \hat{\mathbf{X}}_{\text{HT}}^*)' \mathbf{B}_{\text{NR}} + O_p(n^{-1})$$

und damit ist mit Annahme [A3]

$$\begin{aligned}E[V[\hat{Y}_{\hat{M}^*}^* - \hat{Y}_{\text{HT}}^* - (\hat{\mathbf{X}}^* - \hat{\mathbf{X}}_{\text{HT}}^*)' \hat{\mathbf{B}}_{\text{NR}} | \mathcal{F}] | \mathcal{H}] &= \\ \frac{1}{N^2} \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i} (1 - 1/n_i) (\theta_{kl} - \theta_k \theta_l) (w_k e_k^*) (w_l e_l^*) &+ O_p(n^{-1})\end{aligned} \quad (5.5)$$

wobei $e_k^* = y_k - \mathbf{x}'_k \mathbf{B}_{\text{NR}}$ für $k \in U$ ist. Damit ergibt sich für die Fälle C1, B1, B2 und A1

$$E[V[\hat{Y}_{\hat{M}^*}^* - Z^* | \mathcal{F}] | \mathcal{H}] = V[\hat{Y}_M^* | \mathcal{H}] + O_p(n^{-1})$$

mit

$$\begin{aligned}\text{Fall C1: } Z^* &= \hat{Y}_{\text{HT}}^* \\ \text{Fall B1: } Z^* &= \hat{Y}_{\text{HT}}^* - \hat{\mathbf{X}}_{\text{AHT}}^{*'} \hat{\mathbf{B}}_{\text{NR}}^{\text{A}} \\ \text{Fall B2: } Z^* &= \hat{Y}_{\text{HT}}^* - \hat{\mathbf{X}}_{\text{HT}}^{*'} \hat{\mathbf{B}}_{\text{NR}} \\ \text{Fall A1: } Z^* &= \hat{Y}_{\text{HT}}^* - \hat{\mathbf{X}}_{\text{AHT}}^{*'} \hat{\mathbf{B}}_{\text{NR}}^{\text{A}}\end{aligned}$$

Außerdem ist

$$E[E[\hat{Y}_{\hat{M}^*}^* | \mathcal{F}] | \mathcal{H}] = \hat{Y}_{\text{HT}}^* + (\hat{\mathbf{X}} - \hat{\mathbf{X}}_{\text{HT}})' \mathbf{B}_{\text{NR}} + O_p(n^{-1}) \quad (5.6)$$

Sowohl $E[V[\hat{Y}_{\hat{M}^*}^* - Z^* | \mathcal{F}] | \mathcal{H}]$ als auch $E[E[\hat{Y}_{\hat{M}^*}^* | \mathcal{F}] | \mathcal{H}]$ lassen sich durch Monte-Carlo Simulation hinreichend genau approximieren. Somit wird das folgende Simulations-Verfahren verwendet.

Bootstrap. Ziehe I Stichproben und aus jeder Stichprobe jeweils J Bootstrap-Stichproben. Berechne \hat{Y}_{ij}^* auf Basis von (5.3) und Z_{ij}^* auf Basis von Z^* für $i = 1, 2, \dots, I$ und $j = 1, 2, \dots, J$. Mit

$$Z_{i\cdot}^* = \frac{1}{J} \sum_{j=1}^J Z_{ij}^*, \quad \hat{Y}_{i\cdot}^* = \frac{1}{J} \sum_{j=1}^J \hat{Y}_{ij}^* \quad \text{und} \quad \hat{Y}_{\cdot\cdot}^* = \frac{1}{I} \sum_{i=1}^I \hat{Y}_{i\cdot}^*$$

ist der Bootstrap-Simulations-Schätzer für den Bias

$$\hat{B}[\hat{Y}_{\hat{M}^*}^*] = \hat{Y}_{\cdot\cdot}^* - Y$$

und der Bootstrap-Simulations-Schätzer für die Varianz

$$\hat{V}[\hat{Y}_{M^*}] = \frac{1}{I} \sum_{i=1}^I (\hat{Y}_{i^*} - \hat{Y}_{..})^2 + \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J ((\hat{Y}_{ij^*} - Z_{ij}^*) - (\hat{Y}_{i^*} - Z_{i^*}^*))^2$$

□

Für den PSA-Schätzer (Fall P1) und den PSA-Schätzer mit Kalibration (Fall P2) lässt sich das Verfahren ebenso anwenden. Die Korrekturterme sind jeweils

$$\begin{aligned} \text{Fall P1: } Z^* &= \hat{Y}_{HT}^* \\ \text{Fall P2: } Z^* &= \hat{Y}_{HT}^* - \hat{\mathbf{X}}_{AHT}^{*'} \hat{\mathbf{B}}^A \end{aligned}$$

mit $\hat{\mathbf{B}}^A = (\sum_{U_1} \sum_{U_i} 1_{\{k \in s\}} w_k \mathbf{x}_{Ak} \mathbf{x}'_{Ak})^{-1} \sum_{U_1} \sum_{U_i} 1_{\{k \in s\}} w_k \mathbf{x}_{Ak} y_k$. Die Herleitung für die Fälle P1 und P2 sind im Anhang A.4 dargestellt.

5.4 Ergebnisse der Simulation

Die Tabelle 1 zeigt die zur Simulation verwendeten Eckdaten. Die Eckdaten umfassen die definierten Schichten, die Anzahl der Gebiete je Schicht, die Anzahl der Haushalte und Personen je Schicht, sowie die Größe der Bevölkerung in der jeweiligen Schicht. Da Hamburg, Bremen und das Saarland als eigene Schichten zu klein waren, wurden sie Schleswig-Holstein, Niedersachsen bzw. Rheinland-Pfalz zugeordnet. Bei der Simulation wurden 20% der Gebiete je Schicht gezogen. Für Brandenburg, Mecklenburg-Vorpommern und Sachsen-Anhalt entspricht das 4 gezogenen Gebieten. Die Schicht mit der Anzahl der maximal gezogenen Gebiete ist Bayern (20 Gebiete). Hierbei ist noch zu erwähnen, dass Thüringen im SOEP etwas überrepräsentiert ist, 8% versus 3% in der Bevölkerung. Alle anderen Schichten im SOEP sind proportional zur Verteilung in der Bevölkerung vertreten. Aufgrund der Ziehung von mehr als einem Gebiet je Schicht im SOEP, unterscheidet sich das Ziehungsverfahren von dem des Mikrozensus. Im Mikrozensus wird nur ein Auswahlbezirk je Schicht gezogen. Dies ermöglicht im Allgemeinen keine unverzerrte Schätzung von Varianzen. Rendtel und Schimpl-Neimanns (2001) haben eine mögliche Lösung für dieses Problem vorgeschlagen.

Tabelle 1: Gebiete, Haushalte und Personen je Schicht

Schicht	Gebiete	Haushalte	Personen	Bev. [1000]
Schleswig-Holstein, Hamburg	26	191	423	4'407.3
Niedersachsen, Bremen	48	523	1'276	8'433.1
Nordrhein-Westfalen	78	1'125	2'792	17'830.8
Hessen	30	387	900	5'990.3
Rheinland-Pfalz, Saarland	38	325	821	5'065.9
Baden-Württemberg	52	714	1'865	10'260.1
Bayern	99	744	1'780	11'970.3
Berlin	23	250	532	3'418.4
Brandenburg	19	174	435	2'537.1
Mecklenburg-Vorpommern	18	252	604	1'813.3
Sachsen	24	284	718	4'532.8
Sachsen-Anhalt	19	293	759	2'721.0
Thüringen	34	476	1'185	2'486.4
Gesamt	508	5738	14'090	81'466.8

Die Tabelle 2 enthält Informationen über die 3-Jahres-Mobilität in den betrachteten Schichten.

Zusätzlich ist die räumliche Mobilität getrennt für die Einpersonen- und Mehrpersonen-Haushalte dargestellt. Die mit N beschrifteten Spalten enthalten die Anzahl der Personen in den jeweiligen Schichten, die Anzahl der Personen in Einpersonen-Haushalten und die Anzahl der Personen in Mehrpersonen-Haushalten. Die mit 3JM beschrifteten Spalten enthalten die Anzahl der Personen, die innerhalb der Jahre 1996-1999 räumlich mobil waren. Zum Beispiel sind in Berlin 171 Personen zwischen 1996 und 1999 umgezogen. Von den Umgezogenen lebten im Jahr 1996 25 Personen in einem Einpersonen-Haushalt und 146 Personen in einem Mehrpersonen-Haushalt. Der Anteil der räumlich mobilen Personen ist in den Spalten 3JM[%] dargestellt. Mit 32.1% ist der Anteil der räumlich mobilen Personen in Berlin am höchsten. Wie bereits bei Bašić (2008) gezeigt wurde, ist die räumliche Mobilität von Einpersonen-Haushalten deutlich höher als bei Personen, die in Mehrpersonen-Haushalten leben. Für Berlin, Brandenburg und Thüringen ist der Anteil der mobilen Personen in Mehrpersonen-Haushalten höher als der Anteil der mobilen Personen in Einpersonen-Haushalten.

Tabelle 2: 3-Jahres-Mobilität (3JM) von Personen in Ein- und Mehrpersonen-Haushalten

Schicht				in EinpersonenHHe			in MehrpersonenHHe		
	N	3JM	3JM[%]	N	3JM	3JM[%]	N	3JM	3JM[%]
Schleswig-Holstein, HH	423	109	25.8	75	32	42.7	348	77	22.1
Niedersachsen, HB	1'276	326	25.5	154	46	29.9	1'122	280	25.0
Nordrhein-Westfalen	2'792	672	24.1	300	95	31.7	2'492	577	23.2
Hessen	900	179	19.9	106	40	37.7	794	139	17.5
Rheinland-Pfalz, Saarland	821	219	26.7	77	26	33.8	744	193	25.9
Baden-Württemberg	1'865	478	25.6	193	64	33.2	1'672	414	24.8
Bayern	1'780	396	22.2	211	64	30.3	1'569	332	21.2
Berlin	532	171	32.1	96	25	26.0	436	146	33.5
Brandenburg	435	93	21.4	39	7	17.9	396	86	21.7
Mecklenburg-Vorpommern	604	155	25.7	65	20	30.8	539	135	25.0
Sachsen	718	193	26.9	60	18	30.0	658	175	26.6
Sachsen-Anhalt	759	179	23.6	60	24	40.0	699	155	22.2
Thüringen	1'185	350	29.5	92	27	29.3	1'093	323	29.6
Gesamt	14'090	3'520	25.0	1'528	488	31.9	12'562	3'023	24.1

Die Ergebnisse in Tabelle 3 bestätigen, dass die räumliche Mobilität eine Haushaltsentscheidung ist. Die erste Spalte enthält die Anzahl der Mehrpersonen-Haushalte. Die Anzahl und der Anteil der räumlich immobilien Haushalte (ImmobHHe) sind in den Spalten zwei und drei dargestellt. Die vierte und fünfte Spalte enthalten die Anzahl bzw. den Anteil der Haushalte, deren Haushaltmitglieder alle umgezogen sind (MobHHe). Die sechste und siebte Spalte enthalten die Anzahl bzw. den Anteil der Haushalte, bei denen ein Teil der Haushaltmitglieder ausgezogen ist (SplitHHe). Bei weniger als 10% der Mehrpersonen-Haushalte ziehen einzelne Personen innerhalb des Zeitraums 1996-1999 aus dem gemeinsamen Haushalt aus.

Die Güte der verschiedenen Verfahren wird anhand der Modellierung der Arbeitsmarktdynamik verglichen. Die Arbeitsmarktdynamik umfasst den Wechsel zwischen Erwerbstätigkeit, Erwerbslosigkeit und Nichterwerbstätigkeit. Zur Modellierung der Arbeitsmarktdynamik werden deshalb drei Zustände definiert, erwerbstätig, erwerbslos und Nichterwerbsperson, und die Übergänge zwischen den einzelnen Zuständen geschätzt. Die Tabelle 4 enthält die Verteilung der Veränderung der drei Zustände von 1996 bis 1999. Um in die Analyse einbezogen zu werden, müssen alle betrachteten Personen im Jahr 1996 zwischen 17 und 62 Jahre alt sein. Das sind insgesamt 9'266 (66%) von 14'090 Personen.

Die Schätzung der Übergänge erfolgt anhand der beiden Verfahren, Kalibration und PSA. Die Anpassung der Gewichte bei der Kalibration erfolgt anhand der Merkmale Alter und Erwerbszustand zu

Tabelle 3: Mobilität von Mehrpersonen-Haushalten

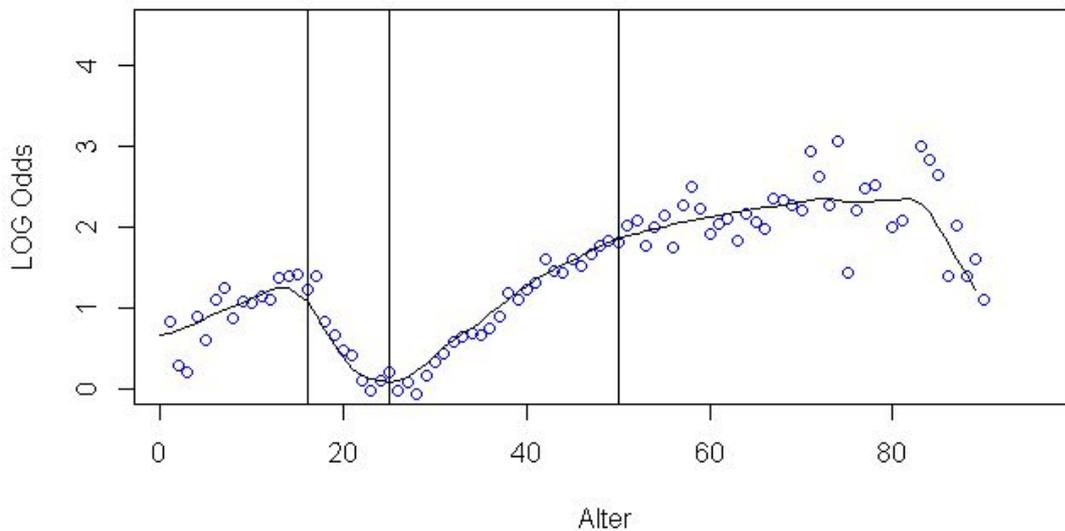
Schicht	N	ImmoHHe		MobHHe		SplitHHe	
		N	[%]	N	[%]	N	[%]
Schleswig-Holstein, Hamburg	116	87	75.0	24	20.7	5	4.3
Niedersachsen, Bremen	369	266	72.1	80	21.7	23	6.2
Nordrhein-Westfalen	825	576	69.8	172	20.8	77	9.4
Hessen	281	220	78.3	41	14.6	20	7.1
Rheinland-Pfalz, Saarland	248	168	67.7	55	22.2	25	10.1
Baden-Württemberg	521	366	70.3	106	20.3	49	9.4
Bayern	533	393	73.7	106	19.9	34	6.4
Berlin	154	94	61.0	47	30.6	13	8.4
Brandenburg	135	99	73.3	24	17.8	12	8.9
Mecklenburg-Vorpommern	187	131	70.0	42	22.5	14	7.5
Sachsen	224	152	67.9	52	23.2	20	8.9
Sachsen-Anhalt	233	173	74.3	42	18.0	18	7.7
Thüringen	384	257	66.9	102	26.6	25	6.5
Gesamt	4'210	2'982	70.8	893	21.2	335	8.0

Tabelle 4: Erwerbsübergänge

		1999		
		ET	EL	NI
1996	Erwerbstätig (ET)	5'814	402	432
	Erwerbslos (EL)	356	244	244
	Nichterwerbspersonen (NI)	479	79	1'216

Beginn des betrachteten Untersuchungszeitraums. Die Anzahl der Personen auf der Populationsebene in den jeweiligen Altersgruppen beträgt: <17 (N=3'287), 17-24 (N=1'278), 25-49 (N=5'768), 50-62 (N=2'220) und >62 (N=1'537). Die Wahl der Altersgruppen orientiert sich an der Arbeit von Bašić (2008). Dort wurde gezeigt, dass die Mobilität durch die Lebensphasen eines Menschen bestimmt wird.

Abbildung 1: Logarithmierte Odds innerhalb 3 Jahren räumlich immobil zu sein.



In den jungen Jahren kommt es zur erhöhten Mobilität aufgrund des Auszuges aus dem Elternhaus

(z.B. Studium). Die zweite Phase (25-49) wird durch Familiengründung und Familienerweiterung beeinflusst. Die räumlich immobilsten sind die Personen in der Altersgruppe ab 49. Im Unterschied zum Alter erfolgt die Kalibrierung bei den Erwerbszuständen auf der Stichprobenebene. Dies ist der Tatsache geschuldet, dass die Information auf der Populationsebene im realen Anwendungsfall nicht bekannt ist. Die künstliche Grundgesamtheit enthält 6'648 erwerbstätige, 844 erwerbslose und 1'774 Nichterwerbspersonen. Für den PSA-Schätzer werden ebenfalls Alter und Erwerbszustände genutzt, wobei das Alter in Form einer stückweisen linearen Funktion berücksichtigt wird.

Die Schätzer werden anhand des relativen Bias miteinander verglichen. Als Benchmark werden zwei Schätzer betrachtet. Der erste Schätzer wird auf Basis des gewichteten Verhältnisses von Stichprobeneinheiten und der Anzahl immobilen Personen in der Stichprobe berechnet:

$$\hat{Y}_{C4} = \frac{\hat{N}_{HT}}{\hat{N}_{NR}} \hat{Y}_{NR}$$

wobei $\hat{N}_{HT} = \sum_U 1_{\{k \in s\}} w_k$, $\hat{N}_{NR} = \sum_U 1_{\{k \in r\}} w_k$ und $\hat{Y}_{NR} = \sum_U 1_{\{k \in r\}} w_k y_k$ ist. Der zweite Schätzer ist definiert als das gewichtete Verhältnis von Stichprobeneinheiten und der kalibrierten Anzahl räumlich immobiler Personen in der Stichprobe, wobei die Kalibrierung anhand der oben definierten Altersgruppen erfolgt:

$$\hat{Y}_{A4} = \frac{\hat{N}_{HT}}{\hat{N}_{NCAL}} \hat{Y}_{NR}$$

mit

$$\begin{aligned} \hat{N}_{NCAL} &= \hat{N}_{HT} + N \hat{\mathbf{B}}_A^{N'} (\mathbf{X}_A - \hat{\mathbf{X}}_{AHT}) \text{ und} \\ \hat{\mathbf{B}}_A^N &= (\sum_U 1_{\{k \in s\}} w_k \mathbf{x}_{Ak} \mathbf{x}'_{Ak})^{-1} \sum_U 1_{\{k \in s\}} w_k \mathbf{x}_{Ak} \end{aligned}$$

Die Tabellen 5, 6 und 7 enthalten die Ergebnisse für die Kalibration. Die ersten drei Spalten beziehen sich auf die Erwerbsübergänge: erwerbstätig nach erwerbstätig, erwerbstätig nach erwerbslos und erwerbstätig zu Nichterwerbsperson. Die mittleren drei Spalten beziehen sich auf die Erwerbsübergänge: erwerbslos nach erwerbstätig, erwerbslos zu bleiben und erwerbslos zu Nichterwerbsperson. Und die letzten drei Spalten auf die Übergänge Nichterwerbsperson nach erwerbstätig, Nichterwerbsperson zu erwerbslos und Nichterwerbsperson zu bleiben. Die 10 Zeilen stellen die Ergebnisse der betrachteten Schätzer dar. C4 ist der Benchmark \hat{Y}_{C4} , C1 berücksichtigt sowohl die Erwerbszustände als auch die Altersgruppen auf Basis der Stichprobe (Horvitz-Thompson Schätzer). Im Gegensatz zu C1 berücksichtigt C3 nur die Erwerbszustände und C2 nur die Altersgruppen. B2 berücksichtigt bei der Kalibration die Altersgruppen auf Populationsebene und B1 die Altersgruppen auf Populationsebene und den Horvitz-Thompson Schätzer der Erwerbszustände auf Stichprobenebene. A4 ist der Benchmark \hat{Y}_{A4} . A1 berücksichtigt die Altersgruppen auf Populationsebene und den Generalised Regression Estimator für die Erwerbszustände. A3 berücksichtigt nur den Generalised Regression Estimator für die Erwerbszustände. A2 ist identisch zu B2, siehe Estevao und Särndal (2002).

Die Ergebnisse basieren auf $I = 1'000$ Stichproben und $J = 500$ Bootstrap-Stichproben je Stichprobe. Sei \hat{Y}_{ij}^* der berechnete Schätzer auf Basis der i -ten Stichprobe und der j -ten Bootstrap-Stichprobe.

Dann ist der geschätzte relative Bias

$$\hat{B}_{\text{REL}}[\hat{Y}_{\text{M}}] = 100 \frac{\hat{Y}_{\text{..}}^* - Y}{Y}$$

wobei $Y = \sum_U y_k$ der Gesamtwert in der künstlichen Population und

$$\hat{Y}_{\text{..}}^* = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \hat{Y}_{ij}^*$$

der Mittelwert über alle 500'000 Simulationen ist.

Tabelle 5: Relativer Bias der geschätzten Erwerbsübergänge in %

	Erwerbstätig (ET) 1996			Erwerbslos (EL) 1996			Nichterwerbsperson (NI) 1996		
	ET 1999	EL 1999	NI 1999	ET 1999	EL 1999	NI 1999	ET 1999	EL 1999	NI 1999
C4	-4.42	1.16	2.38	-13.11	-2.81	9.36	-15.48	-13.73	11.26
C3	-0.90	4.01	7.10	-10.21	-0.55	13.73	-18.76	-17.96	6.37
C2	-0.41	1.02	-0.91	-5.87	-5.19	1.55	-3.79	-8.59	8.22
C1	0.14	1.82	-0.02	-2.58	-1.27	6.11	-7.11	-11.92	3.85
B2	-0.63	0.74	-1.07	-6.07	-5.37	1.26	-3.95	-9.06	7.90
B1	-0.25	1.30	-0.55	-2.82	-1.53	5.51	-6.35	-12.18	4.30
A4	-4.65	0.86	2.16	-13.34	-3.07	9.16	-15.66	-14.06	10.93
A3	-0.80	4.03	7.21	-10.10	-0.65	13.82	-18.76	-17.99	6.30
A2	-0.63	0.74	-1.07	-6.07	-5.37	1.26	-3.95	-9.06	7.90
A1	-0.08	1.53	-0.21	-2.78	-1.42	5.81	-7.29	-12.45	3.56

In den einzelnen Zellen der Tabelle 5 ist der relative Bias des jeweiligen Schätzers für den jeweiligen Übergang ausgewiesen. Die Ergebnisse zeigen, dass es zum Teil große Unterschiede in der Performance der einzelnen Schätzer gibt. Betrachtet man die Übergänge aus der Erwerbstätigkeit, korrigiert der Schätzer A1 erwerbstätig zu bleiben bzw. Nichterwerbsperson zu werden ganz gut, während der Bias für den Erwerbsübergang in die Erwerbslosigkeit leicht erhöht wird. Allerdings war der relative Bias für diesen Übergang sehr klein. Ähnlich gute Performance liefern auch die Schätzer B1 und C1. Dies ist aber nicht verwunderlich, da alle drei Schätzer die Altersgruppen und Erwerbszustände benutzen, die die stärksten Prädiktoren für die betrachteten Übergänge sind. Hierbei ist zu beachten, dass der relative Bias für alle drei Übergänge relativ klein war.

Die Übergänge mit größerem Bias sind die Übergänge aus der Erwerbslosigkeit und die der Nichterwerbspersonen. Bei den Übergängen aus der Erwerbslosigkeit scheinen wiederum die drei Schätzer A1, B1 und C1 den relativen Bias am besten zu korrigieren. Dies ist aber nicht mehr der Fall, wenn man sich die Übergänge der Nichterwerbspersonen anschaut. Dies sind auch die Übergänge mit dem größten relativen Bias. Hierbei geben die Schätzer A2 bzw. B2 und C2 die beste Performance ab. Allerdings bleibt immer noch ein erheblich Teil des Bias unkorrigiert. Eine mögliche Erklärung für die schlechte Performance aller Schätzer bei dem Übergang für Nichterwerbspersonen in die Erwerbslosigkeit ist die geringe Fallzahl (N=79). Das bessere Abschneiden der Schätzer A2/B2 und C2 im Vergleich zu den Schätzern A1, B1 und C1 liegt möglicherweise an der Tatsache, dass der Zustand Nichterwerbsperson aufgrund von linearen Abhängigkeiten bei der Kalibrierung für die Schätzer A1, B1 und C1 nicht berücksichtigt wurde. Beruhigend ist aber zu sehen, dass für Übergänge mit dem größten Bias (erwerbslos zu erwerbstätig, Nichterwerbsperson zu erwerbstätig) eine gute Bias-Korrektur erreicht werden konnte. Vergleicht man die Schätzer, die am besten bei der Bias-Korrektur abschneiden (A1, B1, C1), dann sind durch die leichtere Berechnung C1 und B1 zu bevorzugen, da die initiale

Kalibrierung des Horvitz-Thompson Schätzers entfällt.

In der Tabelle 6 wird analog zu der Tabelle 5 der relative Bias des Standardfehlers gezeigt. Der geschätzte relative Bias des Standardfehlers basiert auf den Ausführungen des letzten Abschnittes. Sei \hat{V}_{ij}^* der berechnete Varianzschätzer auf Basis der i -ten Stichprobe und der j -ten Bootstrap-Stichprobe. Dann ist der geschätzte relative Bias des Standardfehlers

$$\hat{B}_{\text{REL}}[\hat{V}^{1/2}] = 100 \frac{\hat{V}_{..}^{*1/2} - \hat{V}[\hat{Y}_{\hat{M}}]^{1/2}}{\hat{V}[\hat{Y}_{\hat{M}}]^{1/2}}$$

wobei $\hat{V}[\hat{Y}_{\hat{M}}]$ wie im letzten Abschnitt definiert ist und

$$\hat{V}_{..}^* = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \hat{V}_{ij}^*$$

ist. Somit deuten positive Werte in der Tabelle 6 auf eine Überschätzung des Bootstrap-Standardfehlers (Benchmark) hin, während negative Werte auf eine Unterschätzung des Bootstrap-Standardfehlers hindeuten.

Das Ergebnis ist etwas eindeutiger im Vergleich zur Betrachtung des relativen Bias in der Tabelle 5. Schaut man sich die Übergänge mit dem größten Bias (Nichterwerbsperson zu erwerbslos, erwerbslos zu erwerbstätig und Nichterwerbsperson zu erwerbstätig) an, scheint der Schätzer B1, gefolgt vom Schätzer C1 am besten zu performen. Hierbei ist zu beachten, dass der Schätzer B1 die großen Biase sehr gut korrigiert, während es bei kleinen Biasen zur Überkorrektur kommt. Dies ist aber nicht schlimm, da keine scheinbare Genauigkeit suggeriert wird. Das Ergebnis für den Übergang von Nichterwerbsperson zu erwerbslos schneidet trotz der geringen Fallzahl gut ab.

Tabelle 6: Relativer Bias des geschätzten Standardfehlers in %

	Erwerbstätig (ET) 1996			Erwerbslos (EL) 1996			Nichterwerbsperson (NI) 1996		
	ET 1999	EL 1999	NI 1999	ET 1999	EL 1999	NI 1999	ET 1999	EL 1999	NI 1999
C4	3.69	-8.17	-4.15	-9.97	-6.87	-6.68	-8.87	-12.66	-1.31
C3	3.05	-8.11	-5.99	-5.75	-10.50	-3.97	-11.32	-13.56	-3.93
C2	5.74	-9.10	-4.35	-8.93	-7.16	-7.23	-5.58	-12.58	0.17
C1	6.08	-9.15	-4.26	-3.45	-5.08	-5.67	-3.48	-12.82	1.68
B2	-12.11	-11.70	-10.47	-12.74	-13.14	-10.01	-12.78	-16.08	-9.09
B1	13.44	-7.12	-1.05	-2.88	-5.06	-5.38	19.27	-3.88	9.35
A4	-11.45	-10.86	-9.16	-13.39	-12.22	-9.35	-14.84	-15.39	-9.03
A3	-4.52	-9.13	-10.19	-9.79	-13.35	-4.25	-14.20	-15.54	-8.21
A2	-12.11	-11.70	-10.47	-12.74	-13.14	-10.01	-12.78	-16.08	-9.09
A1	-5.59	-11.61	-10.09	-7.20	-11.58	-8.84	-11.49	-16.31	-6.77

Die Performance der Schätzer A1-A4 und B2 für die Übergänge aus der Erwerbstätigkeit ist darauf zurückzuführen, dass die erste Varianzkomponente unterschätzt wird. Die Unterschätzung wird dadurch hervorgerufen, dass in die Berechnung der Varianz das Residuum aus der Kalibration auf die Population einfließt. Dieses Phänomen ist bereits bekannt. Die gängige Vorgehensweise ist eine Korrektur um die verlorenen Freiheitsgrade, vgl. Lundström und Särndal (1999). Hierfür existiert aber keine theoretische Begründung. Außerdem ist für den Schätzer B1 solch eine Korrektur nicht sinnvoll, da es zu einer noch größeren Überschätzung kommen würde. Weiterhin ist zu beachten, dass die Übergänge mit sehr hoher Fallzahl durch den Schätzer für B1 überschätzt werden.

Eine mögliche Erklärung der Unterschätzung könnte der Anteil der Nonresponse-Streuung sein.

Tabelle 7: Relativer Bias des geschätzten Nonresponse-Standardfehlers in %

	Erwerbstätig (ET) 1996			Erwerbslos (EL) 1996			Nichterwerbsperson (NI) 1996		
	ET 1999	EL 1999	NI 1999	ET 1999	EL 1999	NI 1999	ET 1999	EL 1999	NI 1999
C4	-0.29	-2.90	2.12	-19.13	-7.68	15.06	-20.11	-18.50	13.31
C3	3.45	2.10	8.08	-7.56	-3.37	8.20	-25.79	-24.80	6.71
C2	-2.34	-3.33	-1.31	-8.70	-11.41	4.49	-5.01	-12.88	11.65
C1	-1.95	-2.45	-0.39	-4.59	-4.83	3.66	-7.40	-16.64	-7.86
B2	-2.62	-3.72	-1.49	-8.84	-11.42	3.78	-5.04	-13.25	10.77
B1	-2.30	-2.68	-0.63	-4.72	-4.58	3.18	-6.44	-14.69	-6.74
A4	-1.10	-3.38	1.73	-19.31	-8.05	14.43	-20.20	-18.87	12.45
A3	-0.64	2.02	8.00	-7.80	-3.80	7.50	-25.70	-24.76	6.27
A2	-2.62	-3.72	-1.49	-8.84	-11.42	3.78	-5.04	-13.25	10.77
A1	-5.69	-2.95	-0.76	-5.23	-5.40	2.62	-8.01	-16.91	-11.98

Um dies zu untersuchen wurde der relative Bias der einzelnen Schätzer berechnet und in Tabelle 7 dargestellt. Aus der Tabelle 7 ist ersichtlich, dass ein großer relative Bias des Standardfehlers bei den Übergängen aus Erwerbstätigkeit nicht auf die Nonresponse-Streuung zurückzuführen ist. Weiterhin enthüllt die Tabelle 7, dass für die Übergänge erwerbslos zu erwerbstätig und Nichterwerbsperson zu erwerbstätig der relative Bias bei der Verwendung der Schätzer A1, B1 und C1 überproportional korrigiert wird.

Zusammenfassend lässt sich folgende Aussage treffen. Um sowohl für die Übergänge als auch den Standardfehler eine gute Schätzung zu erreichen, ist es notwendig an die Populationsmerkmale und gleichzeitig an die Stichprobeninformation zu kalibrieren. Dass die Erwerbszustände und die Altersgruppen informativ für die Reduktion des Bias ist, wurde im Rahmen des modellbasierten Ansatzes bei Bašić (2008) gezeigt.

Tabelle 8: Relativer Bias des geschätzten Erwerbsübergänge in %, Kalibrationsschätzer und PSA-Schätzer im Vergleich

	Erwerbstätig (ET) 1996			Erwerbslos (EL) 1996			Nichterwerbsperson (NI) 1996		
	ET 1999	EL 1999	NI 1999	ET 1999	EL 1999	NI 1999	ET 1999	EL 1999	NI 1999
C4	-4.42	1.16	2.38	-13.11	-2.81	9.36	-15.48	-13.73	11.26
C1	0.14	1.82	-0.02	-2.58	-1.27	6.11	-7.11	-11.92	3.85
B1	-0.25	1.30	-0.55	-2.82	-1.53	5.51	-6.35	-12.18	4.30
A1	-0.08	1.53	-0.21	-2.78	-1.42	5.81	-7.29	-12.45	3.56
GLM	-0.05	0.00	0.25	1.04	-1.75	3.36	-7.98	-12.69	4.01
GLM+	-0.13	-0.09	0.31	0.93	-1.87	3.34	-7.98	-12.94	3.95
GEE	1.35	3.08	2.51	-5.00	-4.04	2.16	-4.89	-9.63	8.35
GEE+	0.24	0.84	-0.13	-6.22	-6.25	-0.99	-6.58	-12.11	5.29

Die Tabellen 8 und 9 zeigen den relativen Bias der Erwerbsübergänge sowie den relativen Bias der geschätzten Standardfehler für die drei ausgewählten Kalibrationsschätzer (A1, B1 und C1) und die PSA-Schätzer basierend auf dem GLM und GEE-Ansatz. Der Schätzer GLM basiert auf dem Logit-Modell mit stückweiser linearer Funktion des Alters und Erwerbszustände, der Schätzer GLM+ entspricht dem Schätzer GLM erweitert um eine zusätzliche Kalibration (bzgl. des Alters). Die Schätzer GEE und GEE+ enthalten zusätzlich zu den Schätzern GLM und GLM+ die Information über die Korrelation innerhalb der Haushalte.

Für die Übergänge aus der Erwerbstätigkeit in Tabelle 8 wird der Bias bei den Schätzern GLM und GLM+ vollständig korrigiert. Ähnlich gute Performance zeigen die beiden Schätzer für die Übergänge aus der Erwerbslosigkeit. Für die Übergänge mit dem größten Bias (Nichterwerbstätigkeit) ist das Bild

Tabelle 9: Relativer Bias des geschätzten Standardfehlers in %, Kalibrationsschätzer und PSA-Schätzer im Vergleich

	Erwerbstätig (ET) 1996			Erwerbslos (EL) 1996			Nichterwerbsperson (NI) 1996		
	ET 1999	EL 1999	NI 1999	ET 1999	EL 1999	NI 1999	ET 1999	EL 1999	NI 1999
C1	6.08	-9.15	-4.26	-3.45	-5.08	-5.67	-3.48	-12.82	1.68
B1	13.44	-7.12	-1.05	-2.88	-5.06	-5.38	19.27	-3.88	9.35
A1	-5.59	-11.61	-10.09	-7.20	-11.58	-8.84	-11.49	-16.31	-6.77
GLM	0.22	-7.73	-5.09	-7.25	-9.64	-7.41	-6.46	-12.55	-0.34
GLM+	-5.16	-9.82	-8.40	-12.45	-13.94	-10.51	-13.81	-15.86	-3.49
GEE	-0.01	-6.22	-6.03	-9.45	-10.24	-7.60	-8.33	-11.10	-1.33
GEE+	-4.95	-6.96	-6.89	-13.06	-12.72	-7.37	-11.27	-11.80	-2.86

nicht mehr so eindeutig. Hier scheint der GEE-Schätzer bei der Betrachtung aller drei Übergänge die beste Performance abzuliefern.

Tabelle 9 zeigt den relativen Bias für die Standardfehler. Die positiven Werte in der Tabelle 9 deuten auf eine Überschätzung des Bootstrap-Standardfehlers (Benchmark) hin, während negative Werte auf eine Unterschätzung des Bootstrap-Standardfehlers hindeuten. Hier zeigt der GLM Schätzer erneut eine gute Performance, wobei das Ergebnis nicht mehr so eindeutig ist wie bei der Betrachtung der Erwerbsübergänge selbst. Die mittelmäßige Performance der einzelnen Schätzer könnte daran liegen, dass es kein fixes Response-Modell gibt, mit dem die räumliche Mobilität simuliert wird. Bei der nicht-parametrischen Vorgehensweise wird das Response-Modell von Stichprobe zu Stichprobe über die empirische Verteilung jedes mal neu spezifiziert.

6 Zusammenfassung

Das Ziel dieser Arbeit war zu zeigen, ob der Mikrozensus-Längsschnittdatensatz 1996-1999 für die Schätzung von Arbeitsmarktdynamiken in der deutschen Bevölkerung benutzt werden kann. Es wurde gezeigt, dass die Nicht-Berücksichtigung der Ausfälle durch räumliche Mobilität zu verzerrten Schätzungen der Arbeitsmarkübergänge führt. Zur Korrektur dieser Verzerrung wurden zwei Schätzer (Kalibration und PSA) miteinander verglichen. Der Vergleich basierte auf der Betrachtung des relativen Biases sowie der Schätzung des Standardfehlers.

Da es für die beiden Ansätze keine einheitliche Theorie gab, wurden zunächst die beiden Ansätze in einem asymptotischen Kontext dargestellt. Hierbei wurden für beide Verfahren die asymptotisch korrekten Varianzen hergeleitet und ein neues Verfahren zur Simulation der Ausfälle durch Nonresponse entwickelt. Das neue Verfahren wurde auf das SOEP angewendet, das räumliche Mobilität erfasst, um die Performance der beiden Schätzer miteinander zu vergleichen.

Die Ergebnisse der Simulationsstudie zeigten, dass der Kalibrationsansatz bei gleichzeitiger Nutzung von Gesamtwerten der Population und der Stichprobeninformation den Bias bei gleichzeitig leichter Unterschätzung des Standardfehlers reduzieren kann. Zusätzlich zeigte die Simulationsstudie, dass der PSA-Schätzer auf Basis eines einfachen Logit-Modells und ohne Berücksichtigung der Korrelation innerhalb der Haushalte, den Erwerbsübergang erwerbslos zu erwerbstätig noch besser korrigiert. Wird zusätzlich auf den PSA-Schätzer die Kalibration angewendet, ergab sich kein signifikanter Unterschied zum PSA-Schätzer alleine. Bei der Schätzung der Standardfehler zeigte sich, dass bei den beiden Übergängen mit der höchsten Fallzahl der PSA-Schätzer relativ gut abschnitt.

Eine vorsichtige Auswertungsstrategie würde auf beide Verfahren zurückgreifen und bei großen Diskrepanzen auf bedeutende Selektionseffekte durch die Nichterfassung der mobilen Personen schließen.

A Anhang

A.1 Eine untere Schranke für $\hat{\mathbf{T}}_{\text{NR}}$

Die Funktion $f : \mathbb{R}^+ \rightarrow (0, 1]$ definiert durch $f(a) = (1 + a)^{-1}$ ist eine konvex Funktion. Die Gewichte

$$c_k = \frac{1_{\{k \in r\}} w_{\text{Mk}}(\nu' \mathbf{x}_k)^2}{\sum_U 1_{\{k \in r\}} w_{\text{Mk}}(\nu' \mathbf{x}_k)^2}$$

für $\nu \in \mathbb{R}^d$ mit $\|\nu\| = 1$ summieren sich zu Eins:

$$\sum_U c_k = 1$$

Mit der Jensenschen Ungleichung folgt, dass

$$\sum_U c_k f(\lambda'_0 \mathbf{x}_k) \geq f(\sum_U c_k \lambda'_0 \mathbf{x}_k)$$

gilt. Mit der Definition von f ist

$$\sum_U c_k (1 + \lambda'_0 \mathbf{x}_k)^{-1} \geq \frac{1}{\sum_U c_k (1 + \lambda'_0 \mathbf{x}_k)}$$

und somit

$$\sum_U 1_{\{k \in r\}} w_{\text{Mk}}(\nu' \mathbf{x}_k)^2 (1 + \lambda'_0 \mathbf{x}_k)^{-1} \geq \frac{(\sum_U 1_{\{k \in r\}} w_{\text{Mk}}(\nu' \mathbf{x}_k)^2)^2}{\sum_U 1_{\{k \in r\}} w_{\text{Mk}}(\nu' \mathbf{x}_k)^2 (1 + \lambda'_0 \mathbf{x}_k)}$$

Also ist

$$\nu' \hat{\mathbf{T}}_{\text{NR}} \nu \geq \frac{(\nu' \hat{\mathbf{T}}_{\text{M}} \nu)^2}{\sum_U 1_{\{k \in r\}} w_{\text{Mk}}(\nu' \mathbf{x}_k)^2 (1 + \lambda'_0 \mathbf{x}_k)}$$

A.2 Nicht-natürliche Link-Funktionen

Die Matrix $\hat{\mathbf{Q}}$ lässt sich basierend auf der Idee von Fahrmeir und Kaufmann (1985) zerlegen:

$$\hat{\mathbf{Q}}(\lambda) = \hat{\mathbf{Q}}_{\text{A}}(\lambda) + \hat{\mathbf{Q}}_{\text{B}}(\lambda) + \hat{\mathbf{Q}}_{\text{C}}$$

mit

$$\hat{\mathbf{Q}}_{\text{A}}(\lambda) = \sum_U 1_{\{k \in s\}} \{f_{\text{M}}(\lambda, \mathbf{x}_k) - f_{\text{M}}(\lambda_0, \mathbf{x}_k)\} \mathbf{h}_2(\lambda, \mathbf{x}_k)$$

$$\hat{\mathbf{Q}}_{\text{B}}(\lambda) = \sum_U 1_{\{k \in s\}} \{f_{\text{M}}(\lambda_0, \mathbf{x}_k) - 1_{\{k \in r\}}\} \{\mathbf{h}_2(\lambda, \mathbf{x}_k) - \mathbf{h}_2(\lambda_0, \mathbf{x}_k)\}$$

$$\hat{\mathbf{Q}}_{\text{C}} = \sum_U 1_{\{k \in s\}} \{f_{\text{M}}(\lambda_0, \mathbf{x}_k) - 1_{\{k \in r\}}\} \mathbf{h}_2(\lambda_0, \mathbf{x}_k)$$

Sei $A_n \subset A$, $n \in \mathbb{N}$, eine monoton fallende Folge von kompakten Mengen mit $\bigcap_n A_n = \{\lambda_0\}$. Wir wollen zeigen, dass es zu jedem $\varepsilon > 0$ und $\eta > 0$ ein $n_0 \in \mathbb{N}$ gibt, so dass

$$P\left(\left\{n^{-1} \|\hat{\mathbf{Q}}\|_{A_n} > \eta\right\}\right) < \varepsilon$$

für alle $n > n_0$ gilt. Wegen

$$P\left(\left\{n^{-1}\|\hat{\mathbf{Q}}\|_{A_n} > \eta\right\}\right) \leq P\left(\left\{n^{-1}\|\hat{\mathbf{Q}}_A\|_{A_n} > \eta/3\right\}\right) + P\left(\left\{n^{-1}\|\hat{\mathbf{Q}}_B\|_{A_n} > \eta/3\right\}\right) + P\left(\left\{n^{-1}\|\hat{\mathbf{Q}}_C\| > \eta/3\right\}\right)$$

reicht es zu zeigen, dass die Terme auf der rechten Seite gegen Null konvergieren.

Sei $\hat{\mathbf{Q}}_{Cij}$ der (i, j) -te Eintrag von $\hat{\mathbf{Q}}_C$. Die Varianz eines Eintrags $\hat{\mathbf{Q}}_{Cij}$ ist

$$V[\hat{\mathbf{Q}}_{Cij}|\mathcal{F}] = \sum_U 1_{\{k \in s\}} V[1_{\{k \in r\}}|\mathcal{F}](\mathbf{x}_{ik}\mathbf{x}_{jk}f_{L2}(\lambda'_0\mathbf{x}_k))^2$$

da nach Definition von $\mathbf{h}_2(\lambda, \mathbf{x}_k)$ der (i, j) -te Eintrag $\mathbf{h}_{2ij}(\lambda, \mathbf{x}_k) = \mathbf{x}_{ik}\mathbf{x}_{jk}f_{L2}(\lambda'\mathbf{x}_k)$ ist. Wegen $|\mathbf{x}_{ik}\mathbf{x}_{jk}| \leq \mathbf{x}_{ik}^2 + \mathbf{x}_{jk}^2 \leq \mathbf{x}'_k\mathbf{x}_k$ und $V[1_{\{k \in r\}}|\mathcal{F}] = f_{N1}(\lambda_0, \mathbf{x}_k) \leq 1$ ist

$$V[\hat{\mathbf{Q}}_{Cij}|\mathcal{F}] \leq \sum_U 1_{\{k \in s\}} \text{Spur}(\mathbf{h}_2(\lambda_0, \mathbf{x}_k))^2$$

Mit Annahme [R1] ist

$$n^{-1}V[\hat{\mathbf{Q}}_{Cij}|\mathcal{F}] \leq (NM_L)^{-1} \sum_U 1_{\{k \in s\}} w_k \text{Spur}(\mathbf{h}_2(\lambda_0, \mathbf{x}_k))^2$$

und mit Annahme [A1] ist

$$n^{-1}V[\hat{\mathbf{Q}}_{Cij}] \leq (NM_L)^{-1} \sum_U \text{Spur}(\mathbf{h}_2(\lambda_0, \mathbf{x}_k))^2 = O(1)$$

Da $E[\hat{\mathbf{Q}}_C] = 0$ gilt ist $n^{-1}\hat{\mathbf{Q}}_C = O_p(n^{-1/2})$ und somit existiert ein $n_C \in \mathbb{N}$, so dass

$$P\left(\left\{n^{-1}\|\hat{\mathbf{Q}}_C\| > \eta/3\right\}\right) < \varepsilon/3$$

für alle $n > n_C$ gilt.

Für $\hat{\mathbf{Q}}_A$ ist

$$\|\hat{\mathbf{Q}}_A(\lambda)\| \leq \sum_U 1_{\{k \in s\}} |f_N(\lambda, \mathbf{x}_k) - f_N(\lambda_0, \mathbf{x}_k)| \text{Spur}|\mathbf{h}_2(\lambda, \mathbf{x}_k)|$$

und mit Annahme [R1] somit

$$n^{-1}\|\hat{\mathbf{Q}}_A(\lambda)\| \leq (NM_L)^{-1} \sum_U 1_{\{k \in s\}} w_k |f_N(\lambda, \mathbf{x}_k) - f_N(\lambda_0, \mathbf{x}_k)| \text{Spur}|\mathbf{h}_2(\lambda, \mathbf{x}_k)|$$

Die durch $f(\lambda, \mathbf{x}_k) = |f_N(\lambda, \mathbf{x}_k) - f_N(\lambda_0, \mathbf{x}_k)| \text{Spur}|\mathbf{h}_2(\lambda, \mathbf{x}_k)|$ definierte Funktion f hat eine Nullstelle bei λ_0 . Außerdem ist sie stetig. Mit den Annahmen [A1] und [A2] konvergiert $N^{-1} \sum_U 1_{\{k \in s\}} w_k f(\lambda, \mathbf{x}_k)$ auf jeder kompakten Menge in A stochastisch gleichmäßig gegen $N^{-1} \sum_U f(\lambda, \mathbf{x}_k)$ und daher auch die rechte Seite der obigen Ungleichung. Da

$$n^{-1} \sup_{A_n} \|\hat{\mathbf{Q}}_A(\lambda)\| \leq \sup_{A_n} (NM_L)^{-1} \sum_U 1_{\{k \in s\}} w_k f(\lambda, \mathbf{x}_k)$$

gilt, gibt es eine Stichprobengröße $n_A \in \mathbb{N}$, so dass

$$P\left(\left\{n^{-1}\|\hat{\mathbf{Q}}_A\|_{A_n} > \eta/3\right\}\right) < \varepsilon/3$$

für $n > n_A$ gilt.

Für $\hat{\mathbf{Q}}_B$ ist

$$\|\hat{\mathbf{Q}}_B(\lambda)\| \leq \sum_U 1_{\{k \in s\}} \text{Spur} |\mathbf{h}_2(\lambda, \mathbf{x}_k) - \mathbf{h}_2(\lambda_0, \mathbf{x}_k)|$$

und mit Annahme [R1]

$$n^{-1} \|\hat{\mathbf{Q}}_B(\lambda)\| \leq (NM_L)^{-1} \sum_U 1_{\{k \in s\}} w_k \text{Spur} |\mathbf{h}_2(\lambda, \mathbf{x}_k) - \mathbf{h}_2(\lambda_0, \mathbf{x}_k)|$$

Die durch $\text{Spur} |\mathbf{h}_2(\lambda, \mathbf{x}_k) - \mathbf{h}_2(\lambda_0, \mathbf{x}_k)|$ definierte Funktion ist stetig und hat eine Nullstelle bei λ_0 . Mit den Annahmen [A1] und [A2] gibt es daher eine Stichprobengröße $n_B \in \mathbb{N}$, so dass

$$P \left(\left\{ n^{-1} \|\hat{\mathbf{Q}}_B\|_{A_n} > \eta/3 \right\} \right) < \varepsilon/3$$

für $n > n_B$ gilt. Für $n_0 = \max\{n_A, n_B, n_C\}$ folgt damit die obige Behauptung.

Eine alternative Vorgehensweise für $\hat{\mathbf{Q}}_A$ und $\hat{\mathbf{Q}}_B$ ist die folgende. Mit dem Mittelwertsatz ist

$$|f_N(\lambda, \mathbf{x}_k) - f_N(\lambda_0, \mathbf{x}_k)| \leq \|f_{N1}(\cdot, \mathbf{x}_k) \mathbf{h}_1(\cdot, \mathbf{x}_k)\|_{A_1} \|\lambda - \lambda_0\| \leq \|\mathbf{h}_1(\cdot, \mathbf{x}_k)\|_{A_1} \|\lambda - \lambda_0\|$$

für $\lambda \in A_n$ und daher

$$n^{-1} \|\hat{\mathbf{Q}}_A(\lambda)\| \leq (NM_L)^{-1} \sum_U 1_{\{k \in s\}} w_k \|\mathbf{h}_1(\cdot, \mathbf{x}_k)\|_{A_1} \text{Spur} |\mathbf{h}_2(\lambda, \mathbf{x}_k)| M_n$$

wobei M_n der Radius von A_n ist. Falls $M_n = O(n^{-1/2})$ gilt, ist die rechte Seite der Ungleichung mit [A1] $O_p(n^{-1/2})$ und somit auch die linke. Für $\hat{\mathbf{Q}}_B$ ist

$$\text{Spur} |\mathbf{h}_2(\lambda, \mathbf{x}_k) - \mathbf{h}_2(\lambda_0, \mathbf{x}_k)| \leq \sup_{A_n} \|f_{L3}(\lambda' \mathbf{x}_k) \mathbf{x}_k\| \text{Spur} |\mathbf{x}_k \mathbf{x}_k'| M_n$$

wobei f_{L3} die dritte Ableitung von f_L ist. Deshalb ist

$$n^{-1} \|\hat{\mathbf{Q}}_B(\lambda)\| \leq (NM_L)^{-1} \sum_U 1_{\{k \in s\}} w_k \sup_{A_n} \|f_{L3}(\lambda' \mathbf{x}_k) \mathbf{x}_k\| \text{Spur} |\mathbf{x}_k \mathbf{x}_k'| M_n$$

und daher $n^{-1} \|\hat{\mathbf{Q}}_B(\lambda)\| = O_p(n^{-1/2})$. Damit ist sogar

$$n^{-1} \|\hat{\mathbf{Q}}\|_{A_n} = O_p(n^{-1/2})$$

A.3 Approximation der GEE Lösung

Die negativen Ableitungen der Score-Funktion nach λ und α sind

$$\begin{aligned} \hat{\mathbf{H}}_{11}(\lambda, \alpha) &= - \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i, k \neq l} w_{Gkl}(\alpha) (1/2 - f_N(\lambda, \mathbf{x}_k)) \mathbf{h}_1(\lambda, \mathbf{x}_k) \mathbf{x}_k' R_l(\lambda) \\ &\quad + \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i, k \neq l} w_{Gkl}(\alpha) \mathbf{h}_1(\lambda, \mathbf{x}_k) \mathbf{x}_l' (1/2 - f_N(\lambda, \mathbf{x}_l)) R_l(\lambda) \\ &\quad + \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i} w_{Gkl}(\alpha) \mathbf{h}_1(\lambda, \mathbf{x}_k) \mathbf{h}_1(\lambda, \mathbf{x}_l)' \\ \hat{\mathbf{H}}_{12}(\lambda, \alpha) &= - \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i} w_{G1kl}(\alpha) \mathbf{h}_1(\lambda, \mathbf{x}_k) R_l(\lambda) \end{aligned}$$

wobei w_{G1kl} die Ableitungen der Gewichte nach α sind. Die Fisher Information ist der Erwartungswert von $\hat{\mathbf{H}}_{11}$:

$$\hat{\mathbf{F}}(\lambda, \alpha) = \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i} w_{Gkl}(\alpha) \mathbf{h}_1(\lambda, \mathbf{x}_k) \mathbf{h}_1(\lambda, \mathbf{x}_l)'$$

Außerdem sei $\hat{\alpha}_1$ die erste Ableitung von $\hat{\alpha}$ bezüglich des Parameters λ .

Betrachtet man die Zerlegung

$$\mathbf{u}(\hat{\lambda}, \hat{\alpha}(\hat{\lambda})) - \mathbf{u}(\lambda_0, \alpha_0) = (\mathbf{u}(\hat{\lambda}, \hat{\alpha}(\hat{\lambda})) - \mathbf{u}(\lambda_0, \alpha_0)) + (\mathbf{u}(\lambda_0, \hat{\alpha}(\lambda_0)) - \mathbf{u}(\lambda_0, \alpha_0))$$

dann gilt mit dem Mittelwertsatz für den zweiten Teil auf der rechten Seite der Gleichung

$$\mathbf{u}(\lambda_0, \hat{\alpha}(\lambda_0)) - \mathbf{u}(\lambda_0, \alpha_0) = - \int_0^1 \hat{\mathbf{H}}_{12}(\lambda_0, \hat{\alpha}_P(a)) da (\hat{\alpha}(\lambda_0) - \alpha_0) \quad (\text{A.1})$$

wobei $\hat{\alpha}_P(a) = \alpha_0 + a(\hat{\alpha}(\lambda_0) - \alpha_0)$ ist. Das Integral auf der rechten Seite hat den Erwartungswert Null, da $E[\mathbf{R}_l(\lambda_0) | \mathcal{F}] = 0$ gilt. Außerdem ist das Integral

$$F_{kl}(\hat{\alpha}_P) = \int_0^1 w_{G1kl}(\hat{\alpha}_P(a))' da$$

gleichmäßig beschränkt, da $\hat{\alpha}(\lambda_0)$ in einer kompakten Menge liegt, die nicht die Eins enthält. Somit gilt

$$E \left[\left(\frac{\sqrt{n_1}}{M_L N_1} \sum_{U_1} 1_{\{i \in s_1\}} w_{i1} \sum \sum_{U_i} F_{kl}(\hat{\alpha}_P) \mathbf{h}_1(\lambda_0, \mathbf{x}_k) \mathbf{R}_l(\lambda_0) \right)^{2p} \right] < M$$

für ein $M > 0$. Mit Annahme [R1] gilt die Ungleichung

$$\begin{aligned} & \frac{1}{n_1^p} \sum_{U_1} 1_{\{i \in s_1\}} E \left[\left(\sum \sum_{U_i} F_{kl}(\hat{\alpha}_P) \mathbf{h}_1(\lambda_0, \mathbf{x}_k) \mathbf{R}_l(\lambda_0) \right)^{2p} | \mathcal{F} \right] \\ & \leq \frac{n_1^p}{(M_L N_1)^{2p}} \sum_{U_1} 1_{\{i \in s_1\}} E \left[\left(w_{i1} \sum \sum_{U_i} F_{kl}(\hat{\alpha}_P) \mathbf{h}_1(\lambda_0, \mathbf{x}_k) \mathbf{R}_l(\lambda_0) \right)^{2p} | \mathcal{F} \right] \end{aligned}$$

für jedes $p \in \mathbb{N}$. Da nach Voraussetzung Personen aus unterschiedlichen Haushalten unabhängig sind, gilt

$$\begin{aligned} & E \left[\left(\frac{1}{\sqrt{n_1}} \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i} F_{kl}(\hat{\alpha}_P) \mathbf{h}_1(\lambda_0, \mathbf{x}_k) \mathbf{R}_l(\lambda_0) \right)^{2p} \right] \\ & \leq E \left[\left(\frac{\sqrt{n_1}}{M_L N_1} \sum_{U_1} 1_{\{i \in s_1\}} w_{i1} \sum \sum_{U_i} F_{kl}(\hat{\alpha}_P) \mathbf{h}_1(\lambda_0, \mathbf{x}_k) \mathbf{R}_l(\lambda_0) \right)^{2p} \right] \end{aligned}$$

und somit ist das Integral in (A.1) beschränkt. Es folgt

$$\mathbf{u}(\lambda_0, \hat{\alpha}(\lambda_0)) - \mathbf{u}(\lambda_0, \alpha_0) = O_p(1)$$

Betrachten man nun den ersten Teil der obigen Zerlegung, so ist

$$\mathbf{u}(\lambda_0, \hat{\alpha}(\lambda_0)) = \int_0^1 \hat{\mathbf{H}}_{11}(\hat{\lambda}_P(a), \hat{\alpha}(\hat{\lambda}_P(a))) + \hat{\mathbf{H}}_{12}(\hat{\lambda}_P(a), \hat{\alpha}(\hat{\lambda}_P(a))) \hat{\alpha}_1(\hat{\lambda}_P(a)) da (\hat{\lambda} - \lambda_0) \quad (\text{A.2})$$

mit $\hat{\lambda}_P(a) = \lambda_0 + a(\hat{\lambda} - \lambda_0)$. Mit der Zerlegung

$$R_l(\lambda) = R_l(\lambda_0) + R_l(\lambda_0) \frac{\sqrt{f_{N1}(\lambda, \mathbf{x}_k)} - \sqrt{f_{N1}(\lambda_0, \mathbf{x}_k)}}{\sqrt{f_{N1}(\lambda, \mathbf{x}_k)}} + \frac{f_N(\lambda_0, \mathbf{x}_k) - f_N(\lambda, \mathbf{x}_k)}{\sqrt{f_{N1}(\lambda, \mathbf{x}_k)}}$$

sind die mit $\sqrt{n_I}$ normierten Terme auf der rechten Seite von (A.2), die $R_l(\lambda)$ enthalten, im p -ten Mittel beschränkt. Also ist

$$\sqrt{n_I}(\hat{\lambda} - \lambda_0) = \left(n_I^{-1} \int_0^1 \hat{\mathbf{F}}(\hat{\lambda}_P(a)) da \right)^{-1} \left(\frac{1}{\sqrt{n_I}} \mathbf{u}_M + \frac{1}{\sqrt{n_I}} (\mathbf{u}(\lambda_0, \hat{\alpha}(\lambda_0)) - \mathbf{u}_M) + O_p(1)(\hat{\lambda} - \lambda_0) \right) \quad (\text{A.3})$$

wobei $\mathbf{u}_M = \mathbf{u}(\lambda_0, \alpha_0)$ und $O_p(1)$ im p -ten Mittel beschränkt ist. Also ist der Erwartungswert von $n^p \|\hat{\lambda} - \lambda_0\|^{2p}$ beschränkt.

A.4 Bootstrap des PSA-Schätzers

Der Bootstrap-Schätzer für die Korrelation (4.27) auf Basis der Pearson-Residuen (4.28) ist

$$\hat{\alpha}^*(\lambda) = \frac{2}{N_i(N_i - 1)} \sum_{U_I} M_i 1_{\{i \in s_I\}} \sum \sum_{U_i, k < l} R_k(\lambda) R_l(\lambda)$$

Die Score-Funktion auf Basis von (4.29) ist

$$\mathbf{u}^*(\lambda, \alpha) = \sum_{U_I} M_i 1_{\{i \in s_I\}} \sum \sum_{U_i} w_{Gkl}(\alpha) \mathbf{h}_1(\lambda, \mathbf{x}_k) R_l(\lambda)$$

wobei die Gewichte $w_{kl}(\alpha)$ wie in (4.25) und (4.26) definiert sind. Der Schätzer $\hat{\lambda}^* : \Omega \rightarrow A$ ist die Lösung von

$$\mathbf{u}^*(\lambda, \hat{\alpha}^*(\lambda)) = 0$$

Die Bootstrap-Variante der Fisher Information ist

$$\hat{\mathbf{F}}^*(\lambda, \alpha) = \sum_{U_I} M_i 1_{\{i \in s_I\}} \sum \sum_{U_i} w_{Gkl}(\alpha) \mathbf{h}_1(\lambda, \mathbf{x}_k) \mathbf{h}_1(\lambda, \mathbf{x}_l)'$$

Somit gilt mit $\hat{\mathbf{F}}_M^* = \hat{\mathbf{F}}^*(\lambda_0, \alpha_0)$, $\mathbf{u}_M^* = \mathbf{u}^*(\lambda_0, \alpha_0)$ und den Annahmen aus Abschnitt 3

$$\hat{\lambda}^* - \lambda_0 = \hat{\mathbf{F}}_M^{*-1} \mathbf{u}_M^* + O_p(n^{-1})$$

Einsetzen in (5.4) ergibt

$$\hat{Y}_{M^*}^* = \hat{Y}_M^* + \hat{\mathbf{D}}_M^{*'} \hat{\mathbf{F}}_M^{*-1} \mathbf{u}_M^* + O_p(n^{-1})$$

wobei

$$\hat{\mathbf{D}}_M^* = -\frac{1}{N} \sum_{U_1} \sum_{U_i} M_i 1_{\{k \in r \cap s\}} \exp\{-\lambda'_0 \mathbf{x}_k\} \mathbf{x}_k y_k$$

ist. Mit

$$\mathbb{E}[\hat{\mathbf{D}}_M^* | \mathcal{H}] = -\frac{1}{N} \sum_{U_1} 1_{\{i \in s_1\}} \sum_{U_i} \exp\{-\lambda'_0 \mathbf{x}_k\} \mathbf{x}_k y_k$$

ist

$$\hat{Y}_{M^*}^* = \hat{Y}_M^* + \mathbb{E}[\hat{\mathbf{D}}_M^* | \mathcal{H}]' \hat{\mathbf{F}}_M^{-1} \mathbf{u}_M^* + O_p(n^{-1})$$

Da wegen $\mathbb{E}[\mathbf{R}_k(\lambda_0) \mathbf{R}_l(\lambda_0) | \mathcal{H}] = 0$ für $k \in U_i$ und $l \in U_j$, $i \neq j$

$$\mathbb{E}[\mathbf{V}[\mathbf{u}_M^* | \mathcal{F}] | \mathcal{H}] = \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i} (1 - 1/n_1) \omega_{kl}(\alpha_0) \mathbf{h}_1(\lambda_0, \mathbf{x}_k) \mathbf{h}_1(\lambda_0, \mathbf{x}_l)'$$

folgt und

$$\mathbb{E}[\text{Cov}[\hat{Y}_{M^*}^*, \mathbf{u}_M^* | \mathcal{F}] | \mathcal{H}] = \frac{1}{N} \sum_{U_1} 1_{\{i \in s_1\}} \sum_{U_i} (1 - 1/n_1) \exp\{-\lambda'_0 \mathbf{x}_k\} \mathbf{x}_k y_k$$

gilt, ist

$$\begin{aligned} \mathbb{E}[\mathbf{V}[\hat{Y}_{M^*}^* - \hat{Y}_{HT}^* | \mathcal{F}] | \mathcal{H}] &= \frac{1}{N^2} \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i} (1 - 1/n_1) (\theta_{kl} - \theta_k \theta_l) (w_k y_k) (w_l y_l) \\ &\quad - \mathbb{E}[\hat{\mathbf{D}}_M^* | \mathcal{H}]' \sum_{U_1} 1_{\{i \in s_1\}} \sum \sum_{U_i} (1 - 1/n_1) \omega_{kl}(\alpha_0) \mathbf{h}_1(\lambda_0, \mathbf{x}_k) \mathbf{h}_1(\lambda_0, \mathbf{x}_l)' \mathbb{E}[\hat{\mathbf{D}}_M^* | \mathcal{H}] + O_p(n^{-1}) \end{aligned}$$

Der Fall P2 folgt analog.

B Englische Zusammenfassung

The objective of this work was to find out whether the first panel version of the German Microcensus, covering the years 1996-1999, can be used for the estimation of labour force flows. The main concern about the use of the German Microcensus for the estimation of the gross flows refers to the introduction of the non-response bias due to missing information about residential movers. Furthermore, the evaluation of different correction methods with respect to the reduction of the possible non-response bias was performed.

To examine this questions we used a methodology that carefully reflects the special missingness that is of interest here, namely the drop-out of persons due to residential mobility. This was achieved by the use of a similar survey which covers residential mobility. Thus, our approach relies on the observed behaviour of individuals who drop-out of the German Microcensus.

With respect to correction methods, we used calibration for non-response and propensity score adjustment (PSA). To apply both approaches the determination of weights was necessary. Furthermore, the theoretical properties of both approaches were derived. In addition, a new approach for simulation of non-response was developed.

The results indicate that the calibration approach using the population totals and sampling information led to considerable bias reduction. At the same time the standard errors were slightly underestimated. Moreover, the simulation study showed that the PSA method based on logistic regression without taking into account the correlation within the households reached very strong bias reduction for the transition unemployment to employment. When calibration was applied on the PSA estimator no significant difference to the use of the PSA estimator alone was observed. For the estimation of standard errors the results showed good performance of the PSA approach for the transitions with the highest number of cases.

A careful evaluation would use both approaches and in the presence of large discrepancies conclude that considerable selection effects are induced by missing information about residential movers.

Literatur

- Basu, D. (1971). An essay on the logical foundations of survey sampling, part one. In V. P. Godambe und D. A. Sprott (Eds.), *Foundations of Statistical Inference* (pp. 203–242). Toronto: Holt, Rinehart and Winston.
- Bašić, E. (2008). *The Problem of Missing Residential Mobility Information in the German Mikrocensus*. Hamburg: Verlag Dr. Kovač.
- Bethlehem, J. G. (1988). Reduction of nonresponse bias through regression estimation. *Journal of Official Statistics*, 4, 251–260.
- Billingsley, P. (1999). *Convergence of Probability Measures (2nd Edition)*. New York: Wiley.
- Cassel, C. M., Särndal, C.-E., und Wretman, J. H. (1976). Some results on generalized difference estimation and generalizd regression estimation for finite populations. *Biometrika*, 63, 615–620.
- Deville, J.-C. (1988). Estimation linéaire et redressement sur information auxiliaire d'enquêtes par sondage. In A. Monfort und J. J. Laffon (Eds.), *Mélanges économiques: essais en l'honneur de Edmond Malinvaud* (pp. 915–927). Paris: Economica.
- Deville, J.-C. und Särndal, C.-E. (1992). Calibraton estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376–382.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7, 1–26.
- Estevao, V. M. und Särndal, C.-E. (2002). The ten cases of auxililary information for calibration in two-phase sampling. *Journal of Official Statistics*, 18, 233–255.
- Fahrmeir, L. und Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*, 13, 342–368.
- Fitzgerald, J., Gottschalk, P., und Moffitt, R. (1998). An analysis of sample attrition in panel data: The michigan panel study of income dynamics. *The Journal of Human Resources*, 33, 251–299.
- Fuller, W. A. und Isaki, C. T. (1981). Survey design under superpopulation models. In D. Krewski, J. N. K. Rao, und R. Platek (Eds.), *Current Topics in Survey Sampling* (pp. 199–226). New York: Academic Press.
- Herther-Eschweiler, R. (2003). *Längsschnittdaten aus dem Mikrocensus, Basis für neue Analysemöglichkeiten*. Bonn: unpublished manuscript.
- Horvitz, D. G. und Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663–685.
- Isaki, C. T. und Fuller, W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77, 89–96.
- Kallenberg, O. (2001). *Foundations of Modern Probability (2nd Edition)*. New York: Springer-Verlag.

- Kim, J. K. und Kim, J. J. (2007). Nonresponse weighting adjustment using estimated response probability. *Canadian Journal of Statistics*, 35, 501–514.
- Kim, J. K. und Riddles, M. K. (2012). Some theory for propensity-score-adjustment estimators in survey sampling. *Survey Methodology*, 38, 157–165.
- Kroh, M. und Spiess, M. (2006). *Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2005)*. Berlin: DIW.
- Liang, K.-Y. und Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.
- Lundström, S. und Särndal, C.-E. (1999). Calibraton as a standard method for treatment of nonresponse. *Journal of Official Statistics*, 15, 305–327.
- Nelder, J. A. und Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society A*, 135, 370–384.
- Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica*, 59, 1161–1167.
- Neyman, J. (1938). Contribution to the theory of sampling human populations. *Journal of the American Statistical Association*, 33, 101–116.
- Oh, H. L. und Scheuren, F. J. (1983). Weighting adjustment for unit nonresponse. In W. G. Madow, I. Olkin, und D. B. Rubin (Eds.), *Incomplete Data in Sample Surveys*, volume 2 (pp. 143–184). New York: Academic Press.
- Pötscher, B. M. und Prucha, I. R. (1989). A uniform law of large numbers for dependent and heterogeneous data processes. *Econometrica*, 57, 675–683.
- Rendtel, U. und Schimpl-Neimanns, B. (2001). Die Berechnung der Varianz von Populationsschätzern im Scientific Use File des Mikrozensus. *ZUMA Nachrichten*, 48, 85–116.
- Särndal, C.-E. und Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. New York: Wiley.
- Särndal, C.-E. und Swensson, B. (1987). A general view of estimation for two phases of selection with applications to two-phase sampling and nonresponse. *International Statistical Review*, 55, 279–294.
- Särndal, C.-E., Swensson, B., und Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 5, 119–127.
- Spiess, M. (2005). *Derivation of Design Weights: The Case of the German Socio-Economic Panel (SOEP)*. Berlin: DIW.
- Yates, F. und Grundy, P. M. (1953). Selection without replacement from within strata with probability proportional to size. *Journal of the Royal Statistical Society B*, 15, 253–261.