

Anwendung von Messmodellen bei der Analyse der Wirkung von Luftschadstoffen



Bachelorarbeit

zur Erlangung des Grades des Bachelor of Science in Statistik
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig-Maximilians-Universität zu München

Eingereicht von:

Michael Sauer

Referent:

Prof. Dr. Helmut Küchenhoff

München, 02. August 2011

One of the greatest discoveries a man makes,
one of his greatest surprises,
is to find he can do what he was afraid he couldn't.

– *Henry Ford*

Inhaltsverzeichnis

Abbildungsverzeichnis	III
Tabellenverzeichnis	III
1 Luftverschmutzung und ihre Geschichte	1
2 Luftschadstoffe	3
2.1 Allgemeine Informationen	3
2.1.1 Direkte Messungen	6
2.1.2 Indirekte Messungen	7
2.2 Messfehler und Validität	8
2.2.1 Messfehler	8
2.2.2 Validität	9
2.3 Feinstaub	11
3 Beziehung zwischen Außen- /Innen- und persönlicher Feinstaubbe-	
lastung	13
3.1 Studiendesign und Daten	13
3.2 Ergebnisse	14
4 Klassischer und Berkson-Messfehler	18
4.1 Definition der Fehler	18
4.2 Auswirkungen der Fehler	20
5 Regression Calibration	21
5.1 Algorithmus	22
5.2 Beispiel einfache lineare Regression	22
5.3 Schätzung der Parameter	25
5.3.1 Drei Möglichkeiten	25
5.3.2 Wiederholungsmessungen	25
5.4 Anwendungsbereich	28
6 Vergleich Maximum Likelihood, Multiple Imputation und Regression	
Calibration	29
6.1 Das Modell	29
6.2 Maximum Likelihood	30
6.2.1 Likelihood für X beobachtet	30
6.2.2 Likelihood für X nicht beobachtet	31
6.2.3 Likelihood für gesamte Stichprobe	32
6.3 Regression Calibration	33
6.4 Multiple Imputation	34

6.5	Simulation und Vergleich	37
6.5.1	Vorgehen	37
6.5.2	Ergebnisse	39
7	Regression Calibration in der Praxis	42
7.1	Studie von Strand et al.	42
7.1.1	Regression Calibration in der Studie	43
7.1.2	Verschiedene Methoden für die Schätzung	45
7.1.3	Ergebnisse	46
7.2	Studie von Spiegelman	49
7.2.1	Hauptstudie und externe Validierungsstudie	49
7.2.2	Mehrere Surrogate	51
7.2.3	Heteroskedastische Varianz des Messfehlermodells	53
8	Ansätze in der Ökonometrie	55
8.1	Generalized Method of Moments	56
8.2	IV-Schätzung	57
8.2.1	Allgemeines Konzept	57
8.2.2	Two Stage Least Squares	58
8.2.3	Probleme der IV-Schätzung	60
9	Abschließende Bemerkungen	62
	Anhang	65
	Literatur	IV
	Erklärung	VIII

Abbildungsverzeichnis

1	Erklärung der Expositions begriffe	4
2	Wirkungskette Exposition und Dosis	5
3	Messfehler und ihre Wirkung	21
4	Vergleich des Bias mit und ohne Regression Calibration	24
5	Schätzer für β_1	48
6	Verteilung der NO ₂ -Exposition	53

Tabellenverzeichnis

1	Mediane von PM _{2.5} , der Absorptionsrate und der Elemente	15
2	Median der Korrelationen für Exposition persönlich/außen/innen	16
3	Ergebnisse der Regression der persönlichen Exposition	17
4	Übersicht über die fünf verschiedenen Modelle und deren Parameter	38
5	Relativer RMSE von RC und MI zu ML für β_1	40
6	Perzentile für F_{PEX}	47
7	Schätzungen für α_1 und Residuenvarianz	47
8	Surrogate und entsprechende Gewichte	52
9	Ergebnisse für interne Validierungsstudie	66
10	Ergebnisse für externe Validierungsstudie	67

1 Luftverschmutzung und ihre Geschichte

Seit einiger Zeit ist das Thema Luftverschmutzung aus den Nachrichten und Medien nicht mehr wegzudenken. Kaum ein Tag vergeht, an welchem nicht über CO₂-Werte, Treibhauseffekt oder Feinstaubbelastung gesprochen wird. Momentan stehen vor allem der Atomausstieg und generell die Energieversorgung im Mittelpunkt. Oftmals ist jedoch auch von Feinstaub in Großstädten die Rede, insbesondere wenn zulässige Grenzwerte überschritten werden. Zur Feinstaubregulation wurden in vielen deutschen Städten bereits Umweltzonen eingerichtet. Diese dürfen lediglich von Kraftfahrzeugen befahren werden, welche eine gewisse Feinstaubemission nicht überschreiten. Derartige Maßnahmen sollen die Gesundheit der Menschen verbessern, da vermutet wird, dass Feinstaub direkte negative Effekte auf den menschlichen Körper hat.

Doch das Problem der Luftverschmutzung vor allem in Städten ist nicht neu. Es ist ein Phänomen, welches die Menschheit bereits in früheren Jahrhunderten beschäftigte. Im 19. Jahrhundert war das Pferd das beliebteste Transportmittel, sowohl für Ware als auch für Menschen. Städte wie New York und London konnten die Industrialisierung nur auf Grund der Pferde vorantreiben, da sie auf den schnellen Transport der Waren von außerhalb und innerhalb der Stadt angewiesen waren. Die Wertschöpfungskette konnte nur mit Hilfe der Pferde ein derart hohes Niveau erreichen.

Durch das dramatische Bevölkerungswachstum in den USA im 19. Jahrhundert und dem immer steigenden Lebensstandard stieg der Bedarf an Pferden immens. In den Spitzenzeiten um das Jahr 1900 lebten ca. 200.000 Pferde in der Stadt New York (Levitt u. Dubner, 2009, Seite 9). Pro Pferd ist von ca. 24 Pfund Kot (Levitt u. Dubner, 2009, Seite 9) und einer viertel amerikanischen Gallone (ca. 0.95 Liter) Urin (Morris, 2007, Seite 5) pro Tag auszugehen. Dies entspricht knapp 5 Millionen Pfund Pferdemist und 50.000 Gallonen Urin pro Tag. Da städtische Reinigungsdienste diesen Mengen nicht gewachsen waren, blieb der Mist größtenteils auf der Straße liegen. Dies erzeugte einen extremen Gestank in den Städten. Des Weiteren ist getrockneter Kot eine Feinstaubquelle und dieser wiederum Transportmittel für Bakterien, Pilze und Krankheitserreger (Arndt, 2001, Seiten 16–17). Zudem stoßen Pferde Methan aus, welches ein weitaus schädlicheres Treibhausgas als CO₂ ist. Die typischen Treppen der alten New Yorker Gebäude dienten damals

dem Zweck den Pferdemit aus den Häuser fern zu halten und gerade in Regenzeiten dem steigenden Pegel aus Mist zu entkommen (Levitt u. Dubner, 2009, Seiten 9–10).

Die Städte hatten einen Punkt erreicht, an welchem sie nicht ohne und nicht mit dem Pferd überleben konnten. Die Lösung damals war das Automobil. Es wurde als große Umweltrevolution gefeiert. Genau dieses Auto aber wird heute für einen Großteil unserer Luftverschmutzung verantwortlich gemacht. Zum Einen trägt es zur CO₂-Emission bei, zum Anderen produziert es Feinstaub. Generell gibt es natürliche Feinstaubquellen wie Vulkanausbrüche, Bodenerosion, Waldbrände, Pilze und Sporen sowie vom Menschen geschaffene Quellen. Hierzu zählen vor allem Ausstöße von Kraftfahrzeugen, Fernheizwerken, Öfen und Heizungen in Wohnhäusern, Industrieprozessen und im häuslichen Bereich Zigarettenrauch (Lahl u. Steven, 2005, Seite 705). Es wird davon ausgegangen, dass vor allem Feinstaub von menschlichen Quellen auf Grund der kleineren Partikelgröße Herz-Kreislauf- und Lungenerkrankungen hervorruft. Goldberg u. a. (2001) stellten eine positive Korrelation des Feinstaubes mit der Todesrate fest, Ostro u. a. (2000) mit Todesfällen durch Herzversagen, Hoek u. a. (2001) mit Todesfällen durch Atemwegserkrankungen und Norris u. a. (1999) mit der Häufigkeit von Asthmaerkrankungen.

Um den Effekt von Feinstaub auf den menschlichen Körper exakt quantifizieren zu können, sind gute Messungen der jeweiligen Schadstoffe unverzichtbar. Nur eine genaue Messung kann als sinnvolle Datengrundlage für Studien dienen. Um die Feinstaubbelastung zu überwachen wurden in vielen amerikanischen und europäischen Städten bereits feste Außenmessstationen installiert. Die Werte dieser Messstationen können politisch genutzt werden, um Höchstgrenzen oder sonstige Regulationen festzusetzen. Jedoch ist bei Studien des Gesundheitseffekts letztendlich die persönliche Feinstaubbelastung von Interesse. Daher kann die Messung an Außenstellen als Ersatz für die fehlende persönliche Messung benutzt werden oder es wird für einige Studienteilnehmer die persönliche Feinstaubwolke über einen bestimmten Zeitraum hinweg gemessen. Dabei muss beachtet werden, dass alle physikalischen Messgeräte einen gewissen Messfehler aufweisen. Außerdem entsteht ein Messfehler auch, wenn man nur die Außenmesswerte zur Verfügung hat oder wenn man für einen Teil seiner Stichprobe zwar persönliche Messwerte hat, diese aber auf seine gesamte Stichprobe anwenden will.

Daher beschäftigt sich folgende Arbeit mit der Anwendung von Messmodellen.

Zuerst wird ein Überblick über verschiedene Luftschadstoffe und deren Messung mit Schwerpunkt auf Feinstaub gegeben. Anschließend wird die Studie von Janssen u. a. (2005) vorgestellt, welche die gemessenen Werte von Feinstaub an Außenstellen und von persönlichen Messungen in Beziehung setzt. Um besser verstehen zu können, welche Arten von Fehlern auftreten können und welche verschiedenen Auswirkungen diese haben, wird dann der klassische und der Berkson-Messfehler näher erläutert. Um diese Fehler im Modell zu behandeln, wird daraufhin das Konzept der Regression Calibration vorgestellt, welches anschließend mit der Methode der Maximum Likelihood und der Multiple Imputation verglichen wird. Auch diese Methoden dienen dazu Fehler im Modell zu berücksichtigen. Abschließend wird anhand zweier Studien gezeigt, wie Regression Calibration erfolgreich in der Praxis angewendet werden kann.

2 Luftschadstoffe

Für einen besseren Überblick wird zunächst eine kleine Einführung in das Themengebiet der Luftschadstoffe gegeben. Das folgende Kapitel bezieht sich dabei auf Monn (2001), wenn nichts anderes angegeben wird.

2.1 Allgemeine Informationen

Wenn von Luftschadstoffen die Rede ist, dient dies lediglich als Oberbegriff für viele verschiedene Schadstoffe, die sich in der Luft befinden. Diese Schadstoffe sind jedoch in ihrem Aufbau und ihrer Art zum Teil sehr verschieden. Generell lassen sie sich grob in vier Kategorien einteilen. Als erstes ist hier der Feinstaub zu nennen. Dieser wird im Verlauf dieses Kapitels und der gesamten Arbeit noch genauer betrachtet. Weiter werden die Bioaerosole und Allergene in einer Gruppe zusammengefasst. Hierunter fallen zum Beispiel Pollen, Pilz- und Schimmelpilzsporen, Bakterien und Endotoxine. Die dritte Gruppe bilden Gase wie Stickstoffdioxid (NO_2) und die Treibhausgase wie Kohlenstoffdioxid (CO_2), Methan (CH_4) und Distickstoffoxid (N_2O), besser bekannt als Lachgas. Die letzte Gruppe bilden die photochemischen Schadstoffe. Zu diesen werden hauptsächlich Ozon (O_3), Schwefelsäure (H_2SO_4), Salpetersäure (HNO_3) und Peroxyacetylnitrat ($\text{C}_2\text{H}_3\text{NO}_5$), auch PAN genannt, gezählt.

Die Auswirkungen von Luftschadstoffen können sowohl die Umwelt als auch den Menschen direkt betreffen. Im Folgenden wird stets der direkte Effekt auf den Menschen betrachtet. Um die verschiedenen Situationen des Menschen mit Schadstoffen besser beschreiben zu können, ist es notwendig bestimmte Begriffe zu definieren. Einer der wichtigsten Begriffe im Umfeld mit Menschen und Schadstoffen ist die Exposition. Ott (1982) vereinheitlichte mehrere vorherrschende Definitionen und kam zu dem Ergebnis, dass Exposition vorliegt, wenn eine Person zu einem bestimmten Zeitpunkt an einem bestimmten Ort ist und sich der Schadstoff mit einer bestimmten Konzentration zum gleichen Zeitpunkt am selben Ort befindet. Die Person muss also in Kontakt mit dem entsprechenden Schadstoff kommen. Dies beschreibt jedoch nur einen Zeitpunkt. In Realität bleibt diese Exposition aber über einen gewissen Zeitraum bestehen. Wenn man die Dauer der Exposition berücksichtigt, gelangt man zu der integrierten Exposition, welche berechnet werden kann, indem die Konzentration über die Zeit integriert wird. Abbildung 1 macht diesen Zusammenhang deutlich.

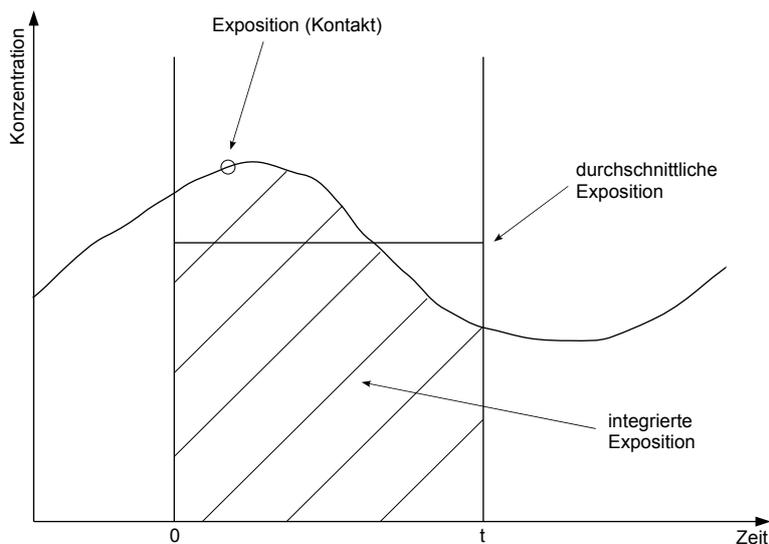


Abbildung 1: Erklärung der Expositionsbegriffe

Jedoch ist es in den meisten Fällen nicht möglich dieses Integral zu berechnen. Daher bedient man sich der durchschnittlichen Exposition. Diese erhält man, indem der Durchschnitt der Schadstoffkonzentration über eine bestimmte Zeitspanne errechnet wird. Insgesamt trifft die Exposition aber nur Aussagen darüber,

welchen Schadstoffen der Mensch gerade ausgesetzt ist, nicht aber darüber, in wie weit diese auf den Körper wirken. So besteht die Möglichkeit, dass ein Schadstoff nicht oder nur zum Teil in den menschlichen Körper eindringt und entsprechend weniger Schaden anrichten kann. Daher wird das Konzept der Dosis eingeführt. Eine Dosis liegt vor, wenn ein Teil des Schadstoffs die physikalischen Grenzen des menschlichen Körpers (z.B. Haut) überwindet (Ott, 1982, Seite 181). Die Dosis, gemessen in Masseneinheiten, bezeichnet also die Menge an Material, welches im Körper absorbiert bzw. abgelagert wird. Zusätzlich wird zwischen der internen bzw. biologisch effektiven Dosis und der potentiellen Dosis unterschieden. Letztere kann zum Beispiel für Feinstaub berechnet werden, indem die integrierte Exposition mit dem Volumen der ausgetauschten Luft in der Lunge pro einer bestimmten Zeitspanne multipliziert wird (Monn, 2001, Seite 2). Abbildung 2 stellt den Zusammenhang zwischen Exposition, Dosis und Gesundheitseffekt dar.¹

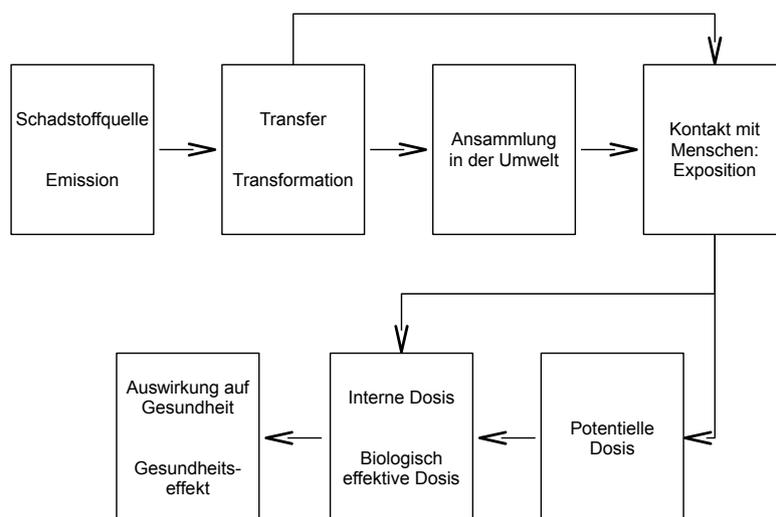


Abbildung 2: Wirkungskette Exposition und Dosis

Um die Dosis oder die Exposition zu bestimmen bzw. zu messen, unterscheidet man generell zwischen direkten und indirekten Messungen. Hierbei gelten die *direkten internen* Messungen als die genauesten in Bezug auf die persönliche Belastung, da Biomarker wie Körperflüssigkeiten oder andere körpereigene Stoffe

¹In Anlehnung an Lioy (1990, Seite 940)

die biologisch effektive Dosis am besten bestimmen können. Ebenfalls sehr gute Ergebnisse liefern die *direkten externen* Messungen. Hierunter versteht man zum Beispiel persönliche Feinstaubmessgeräte, die von Testpersonen über einen längeren Zeitraum direkt am Körper getragen werden. Jedoch ist ein großer Nachteil dieser direkten Messmethoden, dass sie relativ teuer und somit nicht für eine große Stichprobe geeignet sind. Daher bedient man sich oft *indirekten* Messungen. Auch hier gibt es noch weitere Qualitätsstufen. So ist eine stationäre Messung von zum Beispiel Feinstaubpartikeln besser als ein Surrogat wie die Entfernung zur nächsten Strasse. Dies ist jedoch genauer als Daten aus einem qualitativen Fragebogen (z.B. Grad der Verschmutzung: hoch/mittel/niedrig), was wiederum besser ist als qualitative kategoriale Daten (z.B. Exposition: ja/nein) (Monn, 2001, Seite 3).

2.1.1 Direkte Messungen

Bei den direkten Messmethoden wird zwischen den personenbezogenen Probenahmen bzw. Messungen und den biologischen Markern unterschieden. Erstere benötigen zur Messung stets ein externes Gerät, wobei natürlich für verschiedene Schadstoffe unterschiedliche Gerätschaften benutzt werden müssen. Für Gase eignen sich vor allem die sogenannten passiven Sammler, die auch als Diffusions-Sammler bezeichnet werden. Diese beruhen auf dem physikalischen Prozess der thermischen Eigenbewegung von Teilchen. Demnach bewegen sich statistisch gesehen mehr Teilchen aus Bereichen hoher Konzentration in Bereiche niedriger Konzentration, was einen Konzentrationsausgleich bewirkt. Mit Hilfe der Fick'schen Gesetze (Diffusionsgleichung) lässt sich die Konzentration des untersuchten Gases in der Luft dann berechnen (Fick, 1855). Diese Diffusionssammler sind kleine Geräte, die problemlos an der Kleidung befestigt werden können. Zur Messung des Feinstaubes werden oft Gravimeter eingesetzt. Sie messen das zusätzliche Gewicht durch Teilchen, die sich in einem bestimmten Zeitraum auf dem Filter gesammelt haben. Hierbei wird zumeist zwischen $PM_{2,5}$ (Feinstaub mit einer Partikelgröße kleiner als $2.5 \mu m$) und PM_{10} (Feinstaub mit einer Partikelgröße kleiner als $10 \mu m$) unterschieden.

Bei den biologischen Markern wird weitläufig zwischen denen der Exposition und denen des (Gesundheits-)Effekts differenziert. Letztere sind zumeist vorklinische Indikatoren für Anomalien, welche ebenfalls medizinische Diagnosen einschlie-

ßen können. Als Beispiele solcher Marker dienen unter anderem Chromosomen-Anomalien, Zytokine (Glykoproteine, welche das Wachstum und die Differenzierung von Zellen regulieren) oder Änderungen der Lungenfunktion. Die biologischen Marker der Exposition spiegeln die Konzentration des Analyts wider, welches die physikalischen Grenzen des menschlichen Körpers überwunden hat. Biologische Marker können von Atem, Urin, Haaren, Finger- und Fußnägeln, Nasenspülungen, Blut und anderen Spülungen gewonnen werden. Vorteil dabei ist, dass die Exposition über die Zeit integriert ist und alle Wege der Exposition eingeschlossen sind. Dies ist zugleich aber auch ein Nachteil, da nicht mehr zwischen den verschiedenen Wegen der Exposition unterschieden werden kann und Stoffe über die Zeit im Körper auch abgebaut werden. Daher ist zum Beispiel auch das Wissen über die Halbwertszeit der einzelnen Schadstoffe im Körper wichtig.

2.1.2 Indirekte Messungen

Wie oben beschrieben zählen zu den indirekten Messungen alle Methoden, die nicht direkt die Exposition oder Dosis eines Individuums erfassen. Weit verbreitet sind in diesem Bereich Messungen an Außenstationen. In den USA und in Europa wurden Netzwerke etabliert, welche kontinuierlich die Werte mehrere Schadstoffe ermitteln. Zum Teil gibt es mehrere Messstationen pro Stadt in unterschiedlichen Regionen (städtisch/vorstädtisch/ländlich). Diese sind teilweise mit der selben Technik und den selben Gerätschaften ausgestattet wie bei den personenbezogenen Probenahmen, wenn gleich auch größer.

Ein weiterer Ansatz sind sogenannte Mikroumgebungen (ME)². Im täglichen Leben halten sich Personen nicht immer im gleichen Raum oder in der gleichen Umgebung auf. Daher ist der Ansatz der ME's für verschiedene Situationen eigene kleine Umgebungen zu definieren, wobei jeweils eine andere Schadstoffkonzentration und -zusammensetzung angenommen wird. Meistens wird unterschieden in Freiluft (Innenstadt, Wohnumgebung, u.a.), Innenraum (Küche, Wohnzimmer, Schlafzimmer, u.a.), Arbeitsplatz, Transport (Auto, Bus, Bahn, u.a.) und andere Umgebungen (Supermarkt/Einkaufszentren, Restaurant, u.a.). Anhand von Tagebüchern oder Fragebögen lässt sich dann für jede Person zu jedem Zeitpunkt bzw. jeder Zeitspanne eine ME zuordnen. Wird die Konzentration der Schadstoffe

²engl.: microenvironment, abgekürzt mit „ME“

der jeweiligen ME mit dem entsprechenden Zeitraum multipliziert, in welchem sich die Person dort aufgehalten hat, und über einen definierten Zeitraum addiert (Tag/Woche/Monat), erhält man die persönliche integrierte Exposition für jede Testperson.

Zudem existieren physikalische (deterministische) und statistische (stochastische) Modelle, um die Exposition von Personen oder Gruppen vorhersagen zu können. Erstere basieren vor allem auf dem Wissen physikalischer/chemischer Prozesse in der Atmosphäre, wohingegen letztere auf gemessenen Daten und erklärenden Variablen beruhen. Teilweise wird versucht beide Ansätze zu kombinieren und lokale Variablen aufzunehmen, um möglichst präzise Vorhersagen von Schadstoffen in der Luft bzw. der Atmosphäre zu erzeugen.

Als letztes sollen noch die Fragebögen als Mittel der indirekten Messung genannt werden. Sie spielen eine wichtige Rolle, um möglichen Kontakt mit Emissionsquellen zu identifizieren. Dies ist vor allem relevant, um Kontakt mit Quellen im Innenraum (Gasherd, Zigarettenrauch) festzustellen. Des Weiteren werden sie wie oben beschrieben bei Modellen mit ME's benötigt und sind essentiell bei der Analyse von Langzeit-Exposition bei retrospektiven Studien. Dem Nachteil der ungenauen Messung steht der Vorteil der geringen Kosten gegenüber.

Bei all diesen Messungen und Messmethoden muss jedoch berücksichtigt werden, dass Messfehler auftreten können und mit sehr hoher Wahrscheinlichkeit auch auftauchen werden. Daher ist es wichtig zu klären, welche Messfehler existieren und wie die Validität von Messungen überprüft, bzw. bewertet werden kann. Dies wird im Folgenden beschrieben.

2.2 Messfehler und Validität

2.2.1 Messfehler

Zunächst einmal lassen sich die Messfehler in zwei Gruppen einteilen, die systematischen und die zufälligen Messfehler. Die systematischen Messfehler folgen einem bestimmten Muster, bzw. einer Regel. Man kann sich zum Beispiel ein falsch eingestelltes Messgerät vorstellen, welches bei jeder Messung 5 % zuviel misst. Diese Art von Fehler ist nicht weiter kritisch, da kein Bias eintritt. Die Regressionsgerade wird lediglich nach oben oder unter verschoben, die Steigung bleibt jedoch gleich, was bedeutet, dass der Zusammenhang zwischen den Variablen richtig geschätzt

wird.

Mehr Probleme bereitet der zufällige Fehler. Dieser verursacht eine Verzerrung bei der Schätzung des Zusammenhangs. In den meisten Fällen ergibt sich ein Bias Richtung Null, d.h. der Zusammenhang wird schwächer geschätzt als er tatsächlich ist. Jedoch muss auch hier nochmals zwischen dem klassischen und dem Berkson-Fehler differenziert werden. Beim Berksonfehler tritt im Gegensatz zum klassischen Fehler nämlich kein Bias auf. Dieser Unterschied wird ausführlich in Kapitel 4 behandelt.

Weiter lassen sich die Messfehler nochmals unterscheiden in differenzieller und nicht-differenzieller Messfehler. Ersterer liegt vor, wenn das Ausmaß des Fehlers in der Fall- und Kontrollgruppe gleich ist bzw. über einen ganzen Bereich der Exposition nicht variiert. Im Gegensatz dazu ist das Ausmaß des Fehlers bei differenziellen Messfehlern über die Fall- und Kontrollgruppe nicht gleich bzw. verändert sich innerhalb eines Bereichs der Exposition. Beide Fehlerarten erzeugen Bias, weswegen sie bei Modellen und Berechnungen stets berücksichtigt werden müssen.

2.2.2 Validität

Seifert (1995) definiert die Validität einer Messung als den Grad, zu welchem die Messung das misst, was sie vorgibt zu messen. Weiter definiert er drei verschiedene Aspekte von Validität. Die *content validity* beschreibt das Ausmaß, zu welchem die Messung den ganzen Umfang des untersuchten Ereignisses mit einarbeitet. Dies bedeutet, dass content validity gegeben ist, wenn alle Orte bzw. Aktivitäten berücksichtigt wurden, die zur Exposition beitragen könnten. Die *construct validity* bezeichnet das Ausmaß, zu welchem die Messung sich mit dem theoretischen Konzept bezüglich des untersuchten Ereignisses deckt. Diese Validität ist jedoch recht komplex, da teilweise das Konstrukt/das Konzept selbst nicht einfach definiert werden kann.

Als letztes beschreibt die *criterion validity* das Ausmaß, zu welchem die Messung mit dem untersuchten Ereignis korreliert. Sie kann bestimmt werden, indem Werte für die persönliche Exposition mit Werten von Surrogaten verglichen werden. Dies wird als Validitätskoeffizient ρ^2 bezeichnet. Auch wenn es bei Studien immer wünschenswert ist, alle drei Aspekte der Validität zu überprüfen, ist der

Validitätskoeffizient die einzig berechenbare Größe. Gilt $\rho^2 = 1$, so entsteht keine Dämpfung des Effekts und es liegt eine perfekte Messung vor. Dies wird jedoch in Realität nur in den seltensten Fällen erreicht. Bezeichne X den wahren Wert und W den gemessenen Wert, bzw. das Surrogat, so kann der Validitätskoeffizient geschrieben werden als

$$\rho_{XW}^2 = 1 - \frac{\sigma_E^2}{\sigma_W^2} = \frac{\sigma_X^2}{\sigma_W^2} \quad ,$$

wobei σ_W^2 die totale Varianz bezeichnet, σ_X^2 die wahre Fehlervarianz (Zwischen-Subjekt-Varianz) und σ_E^2 die Subjekt-Fehlervarianz (Inner-Subjekt-Varianz). Es wird also berechnet, wie viel der totalen Varianz durch die wahre Fehlervarianz erklärt werden kann.

Liegen in einer Studie mehrere Messungen für jede Person vor, kann der Validitätskoeffizient berechnet werden als

$$\rho_{XW}^2 = \frac{1}{1 + \sigma_E^2/\sigma_X^2} = \frac{1}{1 + \lambda} \quad .$$

Dabei bezeichnet λ die relative Präzision einer Messung, da es das Verhältnis zwischen der Varianz eines Subjekts relativ zur Varianz der wahren Exposition in der Studienpopulation beschreibt.

Der Validitätskoeffizient wird zudem bei der Optimierung des Studiendesigns eingesetzt. Denn meistens ist es das Ziel möglichst geringe Varianzen für den Effektschätzer zu erhalten. Diese lässt sich berechnen als

$$\text{Var}(\beta) = \frac{\text{Var}(\text{error})}{N\rho^2\text{Var}(\text{exp})} \quad .$$

Eine geringe Varianz von β erhält man also bei einer geringen Fehlervarianz des Modells ($\text{Var}(\text{error})$), einer großen Stichprobenzahl N , eines großen Validitätskoeffizienten ρ^2 oder einer großen Varianz der Exposition ($\text{Var}(\text{exp})$) (Bandbreite der Expositionskonzentration zwischen den Studienorten).

Messfehler und Validität müssen bei praktisch jeder Studie mit Messungen berücksichtigt bzw. untersucht werden. Denn nur so können verlässliche Ergebnisse in den Modellen gewährleistet werden, da im Allgemeinen davon auszugehen ist, dass keine perfekten Messgeräte existieren.

2.3 Feinstaub

Im Folgenden wird der Feinstaub als Luftschadstoff mit seinen Definitionen und Eigenschaften genauer betrachtet. Wie oben beschrieben wird er in die Gruppen PM_{10} für Partikel kleiner als $10\ \mu\text{m}$ und $PM_{2.5}$ für Partikel kleiner als $2.5\ \mu\text{m}$ eingeteilt. Dies hat den Grund, dass größere Partikel in dem Bereich $8 - 15\ \mu\text{m}$ zum Großteil von natürlichen Quellen stammen und Partikel in dem Bereich $1 - 2\ \mu\text{m}$ fast ausschließlich vom Menschen verursacht werden. Dadurch wird auch die bimodale Verteilung des Feinstaubes mit je einem Peak bei diesen Partikelgrößen erklärt. In einigen Studien (Clayton u. a., 1993; Dockery u. a., 1993; US-EPA, 1996) wurde das Verhältnis von $PM_{2.5}$ zu PM_{10} mit ca. 0.3 bis 0.7 geschätzt.

Generell kann man davon ausgehen, dass in Städten ein gewisses Feinstaub-Hintergrundniveau existiert. Gerade die Partikel der Größe um $0.5 - 2\ \mu\text{m}$ sind auf horizontaler und niedriger vertikaler (0–100 Meter) Ebene recht gleichverteilt, was zu einer ausgeglichenen Feinstaubbelastung der gesamten betrachteten Region führt. Etwas anders sieht es bei den sehr kleinen (kleiner $0.1\ \mu\text{m}$) und den großen (größer $2.5\ \mu\text{m}$) Partikeln aus. Auf Grund von Koagulation³ bei kleinen Teilchen bzw. der höheren Sinkgeschwindigkeit bei größeren Teilchen ist die räumliche Variation bei diesen Partikeln größer. Für Studien scheint es demnach sinnvoll und kosteneffizient zu sein lediglich eine Messstation für $PM_{2.5}$ zu installieren, aber mehrere Messstellen für PM_{10} und ultrafeine Partikel einzurichten, um eine bessere Differenzierung zu gewährleisten. Eine eigene Messung der ultrafeinen Partikel ist notwendig, da nur eine geringe Korrelation zwischen $PM_{10}/PM_{2.5}$ und den ultrafeinen Partikeln festgestellt werden konnte (Keywood u. a., 1999).

Natürlich bewegen sich Menschen nicht nur außerhalb des Hauses, weswegen auch der Feinstaub innerhalb geschlossener Räume wichtig ist. Dabei ist klar, dass die Feinstaubkonzentration innerhalb des Hauses von der außerhalb beeinflusst wird, da es verhältnismäßig viele Außenquellen gibt. Generell können innen/außen Verhältnisse zwischen ca. 0.2 und 3.5 beobachtet werden. Hier spielen zwei Faktoren eine wichtige Rolle. Zum Einen ist die Art der Ventilation von Bedeutung. Werden Klimaanlage mit guten Filtern verwendet ist das Verhältnis wesentlich geringer als wenn lediglich geöffnete Fenster zur Belüftung dienen. Zum Anderen

³Koagulation (lateinisch *coagulatio* – die Zusammenballung) beschreibt das Vergrößern feiner Teilchen und damit das Aufheben von fein verteilten Zuständen

beeinflussen mögliche Innenquellen dieses Verhältnis. Die wichtigsten sind hierbei Zigarettenrauch (mit Abstand die wichtigste), Gasherd, Holzofen/-kamin und Abstauben bzw. Staubsaugen. In Haushalten mit natürlicher Ventilation und ohne Zigarettenrauch liegen die innen/außen Verhältnisse im Schnitt bei ca. 1, bei Haushalten mit Zigarettenrauch wesentlich höher.

Dennoch ist für den Feinstaub innerhalb des Hauses der Feinstaub von außen die wichtigste Quelle, neben dem Zigarettenrauch. Oezkaynak u. a. (1996) konnten zeigen, dass der Außenfeinstaub in Häusern ohne Zigarettenrauch ca. 75 % des $PM_{2.5}$ und ca. 66 % des PM_{10} innen ausmacht. In Haushalten mit Zigarettenrauch liegen diese Werte noch bei 55–60 %. Zusätzlich spielt die menschliche Aktivität eine Rolle. Viel Bewegung im Haus bedeutet, dass bereits niedergefallene Partikel aufgewirbelt werden und somit zusätzlich in der Luft als Schadstoffe vorhanden sind.

Im Endeffekt ist bei allen Studien aber die persönliche Schadstoffbelastung jeder Person von Interesse. Daher gilt es gut vorhersagbare Beziehungen zwischen dem Außen- /Innenfeinstaub und der persönlichen Exposition herzustellen. Generell scheint eine recht starke Korrelation zwischen dem Außenfeinstaub und der persönlichen Exposition zu bestehen. Janssen u. a. (1998) finden bei day-to-day Studien eine Korrelation von 0.71, welche wesentlich besser ist als die bei Querschnittsstudien. Wilson u. Spengler (1996) konnten feststellen, dass die Beziehung zwischen der an Außenstellen gemessenen Feinstaubbelastung und der persönlichen Exposition umso besser ist, je feiner die Partikel sind. Jedoch müssen individuelle Charakteristika wie zum Beispiel Rauchen und Wohnort an einer stark befahrenen Strasse berücksichtigt werden. So erklärt ein Wohnort an einer großen Strasse und die Zeit, die im Verkehr verbracht wird laut Janssen u. a. (1998) ca. 75 % der Varianz zwischen der persönlichen Belastung und der an Außenstellen gemessenen. Zudem scheint die persönliche Exposition zu PM_{10} generell höher zu sein als die außen bzw. innen vorherrschenden PM_{10} Konzentrationen. Dies könnte gemäß Oezkaynak u. a. (1996) auf die sogenannte persönliche Wolke zurückzuführen sein, welche zum Beispiel aufgewirbelte Partikel von Außenquellen und Partikel an der Kleidung beinhaltet. Eine genauere Analyse der Beziehung zwischen an Außenstellen gemessenem Feinstaub und der persönlichen Belastung wird in Kapitel 3 durchgeführt.

Für die anderen Luftschadstoffe wird keine ausführliche Beschreibung gegeben,

da sich diese Arbeit im Folgenden hauptsächlich am Feinstaub orientieren wird. Es sei jedoch gesagt, dass sich die Bioaerosole und Allergene sehr ähnlich zum Feinstaub verhalten, da sie generell die gleichen Eigenschaften besitzen, jedoch lebende Organismen enthalten. Stickstoffdioxid weist eine recht große räumliche Variation auf, abhängig von der photochemischen Aktivität. Die persönliche Exposition ist besser mit der innerhäuslichen Konzentration korreliert, wobei Gasöfen die größte Quelle im Haus darstellen. Zu guter Letzt sind bei Ozon (Gruppe der photochemischen Schadstoffe) die innerhäuslichen Werte signifikant geringer als die Außenwerte. Dies legt die Einbeziehung individueller Verhaltensweisen bei der Berechnung der persönlichen Exposition nahe.

3 Beziehung zwischen Außen- /Innen- und persönlicher Feinstaubbelastung

Wie im vorherigen Kapitel geschrieben, wird Feinstaub oft an Außenstationen gemessen. Diesbezüglich wurden bereits in vielen europäischen und amerikanischen Städten Außenmessstellen installiert. Kann nun eine vorhersagbare Beziehung zwischen diesen Messstellen und der persönlichen Belastung hergestellt werden, kann auf sehr einfache und kostengünstige Weise die persönliche Exposition ermittelt werden. Denn letztendlich ist in den meisten Studien die persönliche Belastung von Interesse, da diese als potentielle Dosis in den Körper aufgenommen wird und Krankheiten auslöst. Das folgende Kapitel analysiert die Beziehung zwischen dem Feinstaub außerhalb des Hauses, dem innerhalb der Wohnräume und der persönlichen Belastung und bezieht sich auf die Studie von Janssen u. a. (2005), wenn nichts anderes angegeben wird.

3.1 Studiendesign und Daten

Die zugrunde liegenden Daten wurden zwischen 1998 und 1999 in Amsterdam und Helsinki erhoben. Über den Zeitraum eines halben Jahres wurde bei 84 nicht rauchenden Studienteilnehmern alle zwei Wochen eine 24 Stunden dauernde persönliche Messung des Feinstaubes und seiner Bestandteile durchgeführt. Zudem wurde bei jedem Teilnehmer die Feinstaubbelastung im Wohnzimmer an Messge-

räten aufgezeichnet, die in einem Meter Höhe installiert wurden. Hinzu kommen tägliche Messungen des Feinstaubes an fixen Außenstellen, die das städtische Hintergrundniveau widerspiegeln. Die Messgeräte waren der Art, dass nicht nur $\text{PM}_{2.5}$ gemessen, sondern ebenso die einzelnen chemischen Elemente identifiziert werden konnten. Des Weiteren wurde mit Hilfe der Reflexion von den Teilchen der Absorptionsgrad der $\text{PM}_{2.5}$ -Filter untersucht, welcher als Marker für Dieselabgase dient.

Janssen u. a. (2005) empfiehlt für derartige Studien generell ein longitudinales Design. Denn nur eine solche Herangehensweise liefert die für Zeitreihenanalysen relevanten Korrelationen zwischen den Messwerten an Außenstellen und der persönlichen Belastung. Bei Querschnittsdaten werden im Allgemeinen geringere Korrelationen festgestellt. Zudem wird ein größeres Verhältnis von persönlicher Feinstaubexposition zu Außenkonzentrationen beobachtet.

Um verstehen und analysieren zu können wie sich der Feinstaub in der Außen- und Innenluft sowie der persönlichen Belastung zusammensetzt, wurden die einzelnen Elemente des Feinstaubes betrachtet. Von größerem Interesse ist dabei unter anderem Schwefel (S), da dieser kaum häusliche Quellen besitzt und daher die Schwefelpartikel stets von außen kommen, egal wo sie gemessen wurden. Genauso wie Zink (Zn) und Eisen (Fe) zählt Schwefel eher zu den kleinen und leichten Teilchen. Bei diesen wird vermutet, dass sie schnell und einfach von außen in den innerhäuslichen Bereich eindringen und somit auch die persönliche Belastung sehr ähnlich zu den an Außenstellen gemessenen Werten ist. Bei schwereren Partikeln wie Kalium (K), Calcium (Ca), Kupfer (Cu), Silizium (Si) und Chlor (Cl) ist die Vermutung, dass sich diese eher langsamer verteilen und nicht so weit von den Quellen (auch einige Innenquellen) entfernen. Somit fällt die Korrelation zwischen der in der Außenluft gemessenen Konzentration und der persönlichen Belastung geringer aus.

3.2 Ergebnisse

Tabelle 1 gibt einen ersten Überblick über die Daten. Dort sind die Mediane von $\text{PM}_{2.5}$, der Absorptionsrate und der einzelnen Elemente getrennt für beide Städte abgetragen. $\text{PM}_{2.5}$ wird standardmäßig in $\mu\text{g}/\text{m}^3$ gemessen, die Absorptionsrate in $\text{m}^{-1} \times 10^{-5}$ und die Elemente in ng/m^3 . Man kann erkennen, dass die Werte für

	Amsterdam			Helsinki		
	persönlich	innen	außen	persönlich	innen	außen
PM _{2.5}	14.5	14.1	15.7	9.4	9.8	11.4
Abs	1.4	1.5	1.6	1.3	1.4	1.9
S	912.3	1100.6	1299.9	605.3	811.6	1435.7
Zn	13.2	15.7	18.3	11.7	12.4	18.6
Fe	57.0	58.5	71.3	41.6	42.5	79.2
K	87.4	95.3	70.3	103.1	105.2	93.9
Ca	72.9	65.9	40.2	68.5	45.3	36.4
Cu	5.4	8.1	2.5	4.3	5.5	1.8
Si	29.7	55.2	13.7	79.5	74.8	93.9
Cl	40.8	59.2	72.7	9.8	31.9	44.2

Tabelle 1: Mediane von PM_{2.5}, der Absorptionsrate und der Elemente

beide Städte sehr ähnlich sind. Natürlich gibt es eine gewisse Variation, dennoch scheint die Feinstaubbelastung in entwickelten europäischen Großstädten ähnliche Ausmaße anzunehmen. Für PM_{2.5}, die Absorptionsrate und die leichteren Elemente liegt die Exposition außen oberhalb der persönlichen. Entsprechend nimmt das Verhältnis der persönlichen Exposition zur Exposition außen Werte kleiner als Eins an. Für die schwereren Partikel mit verhältnismäßig vielen innerhäuslichen Quellen liegt die persönliche Belastung meist über der Belastung außen mit der Ausnahme von Chlor. Daher sind die Verhältnisse persönlich/außen größer als Eins. Auch bei den Verhältnissen sind beide Städte kaum verschieden. Helsinki weist jedoch tendenziell etwas geringere Quotienten auf.

Für das Verhältnis innen/außen stellen sich fast die gleichen Ergebnisse ein wie für persönlich/außen. Interessant ist hingegen, dass die Verhältnisse persönlich/innen alle samt näher an Eins liegen. Dies ist vermutlich darauf zurückzuführen, dass die Menschen einen größeren Teil ihrer Zeit innen verbringen und nur einen Bruchteil außen. Janssen u. a. (2005, Seite 876) gaben an, dass die Studienteilnehmer insgesamt 88 % der Zeit in geschlossenen Räumen verbrachten. Daher gleicht sich dann die persönliche Belastung stark an die im Haus an.

Von großem Interesse ist aber natürlich auch die Korrelation zwischen den verschiedenen Messstellen (persönlich/innen/außen). Daher werden die Mediane der Spearman-Korrelation in Tabelle 2 zusammengefasst. Wie man erkennen kann, sind die Korrelationen persönlich/außen, innen/außen und persönlich/innen für

	Amsterdam			Helsinki		
	persönlich/ außen	innen/ außen	persönlich/ innen	persönlich/ außen	innen/ außen	persönlich/ innen
PM _{2.5}	0.76	0.79	0.84	0.74	0.80	0.90
Abs	0.91	0.94	0.94	0.89	0.89	0.90
S	0.96	0.96	1.00	0.90	0.90	0.95
Zn	0.89	0.94	0.93	0.70	0.77	0.86
Fe	0.71	0.89	0.80	0.70	0.70	0.71
K	0.70	0.77	0.86	0.55	0.70	0.90
Ca	0.52	0.60	0.64	0.30	0.60	0.70
Cu	0.27	0.21	0.67	0.25	0.10	0.57
Si	0.36	0.26	0.49	0.70	0.71	0.89
Cl	0.52	0.60	0.81	0.30	0.30	0.80

Tabelle 2: Median der Korrelationen für Exposition persönlich/außen/innen

PM_{2.5}, die Absorptionsrate und die leichten Elemente in beiden Städten sehr hoch mit Werten zwischen 0.7 (für Fe) und 1.0 (für S). Dies deutet daraufhin, dass sich diese Partikel in der Luft gut durchmischen und gleichmäßig verteilen. Somit ist die persönliche Exposition annähernd gleich, egal ob sich der Studienteilnehmer innerhalb oder außerhalb des Hauses aufgehalten hat. Bei den schwereren Partikeln waren vor allem die Korrelationen persönlich/außen und innen/außen deutlich geringer als bei den leichteren und auch geringer als die Korrelationen persönlich/innen. Dies liegt an der verhältnismäßig großen Zahl an Innenquellen und der langen im Haus verbrachten Zeit.

Um einen Eindruck für den unabhängigen Einfluss der Konzentration in der Außen- bzw. Innenluft auf die persönliche Exposition zu bekommen, haben Janssen u. a. (2005) folgendes Regressionsmodell geschätzt

$$X_{\text{persönlich}, it} = \alpha_0 + \alpha_1 X_{\text{ausßen}, t} + \alpha_2 X_{\text{innen}, it} + u_{it} \quad .$$

Die Ergebnisse sind in Tabelle 3 zusammengefasst. Dabei bezeichnet *Amsterdam 1* die Situation, wenn alle Daten aus Amsterdam benutzt wurden und *Amsterdam 2* die Situation, wenn alle Daten ausgeschlossen wurden, bei welchen Studienteilnehmer Zigarettenrauch ausgesetzt waren. Die Signifikanzen werden mit Sternen angedeutet (* für $p < 0.10$, ** für $p < 0.05$, *** für $p < 0.01$).

Es fällt auf, dass die Schätzer der Konzentration in der Innenluft stets signifi-

	Amsterdam 1		Amsterdam 2		Helsinki	
	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
PM _{2.5}	0.13*	0.54***	0.06***	0.81***	-0.02	0.97***
Abs	0.07	1.04***	0.25***	0.77***	0.06**	0.98***
S	0.18***	0.59***	0.24***	0.51***	0.05**	0.66***
Zn	0.27***	0.39***	0.31***	0.35***	0.09***	0.81***
Fe	0.07	0.78**	0.04	0.81***	0.04	0.90***
K	-0.29	0.87***	0.15**	0.62***	0.17***	0.57***
Ca	0.23	0.78***	0.22	0.77**	0.23***	0.84***
Cu	0.06	0.56***	0.06	0.57***	-0.12	0.79***
Si	-0.22	1.17***	0.03	0.81***	0.15***	0.44***
Cl	-0.03	0.95***	0.01	0.97***	-0.01	0.37***

Tabelle 3: Ergebnisse der Regression der persönlichen Exposition

kant sind. Auch dies ist wieder darauf zurückzuführen, dass die Studienteilnehmer den Großteil ihrer Zeit im Haus verbrachten. Auf den ersten Blick kontraintuitiv scheint, dass die Schätzer zur Außenluft nur für einige Elemente signifikant sind. Jedoch muss man sich klar machen, dass hier nur der direkte Einfluss der Außenluft auf die persönliche Exposition gemessen wird. Nicht berücksichtigt wird in diesem Modell der Einfluss, den die Außenluft auf die Innenluft hat und somit indirekt auf die persönliche Exposition. Dennoch hat auch die Außenluft bei PM_{2.5}, der Absorptionsrate und den Elementen S, Zn und K (in Helsinki zusätzlich Ca und Si) einen signifikanten direkten Einfluss auf die persönliche Belastung.

Auch zahlenmäßig sind die Ergebnisse relativ hoch. Zum Beispiel führt ein Anstieg der PM_{2.5}-Konzentration der Innenluft um 10 $\mu\text{g}/\text{m}^3$ in Helsinki zu einem Anstieg der persönlichen Exposition um 9.7 $\mu\text{g}/\text{m}^3$. In diesem Fall wird also beinahe der gesamte Anstieg in die persönliche Belastung aufgenommen.

In der Studie von Janssen u. a. (2005) konnten teilweise sehr hohe Korrelationen zwischen der persönlichen Exposition und den an Außenstellen gemessenen Werten bzw. der Messungen im Haus festgestellt werden. Des Weiteren können mit Hilfe der Außen- und Innenmesswerte im Rahmen eines einfachen Regressionsmodells gute Schätzungen für die persönliche Belastung durchgeführt werden. Dies hat zur Folge, dass selbst dann aussagekräftige Ergebnisse in Studien erzielt werden können, wenn lediglich Feinstaubmessungen von Außenmessstellen vorhanden sind. Zusätzlich sind die Ergebnisse für beide Städte sehr ähnlich, was bedeutet,

dass die Resultate für westliche Städte bzw. Regionen gleichermaßen angewendet werden können. Dennoch muss immer berücksichtigt werden, dass Messfehler auftreten können. Daher wird das folgende Kapitel klären welche Arten von Messfehlern existieren und wie sich diese auf die Modelle und Ergebnisse auswirken.

4 Klassischer und Berkson–Messfehler

Nachdem geklärt ist, welche Luftschadstoffe es gibt und wie diese gemessen werden können, wird nun genauer auf die Messfehlerproblematik eingegangen. Nicht nur solche Geräte wie in der Studie aus Kapitel 3 erzeugen einen Messfehler. Auch entstehen Fehler auf Grund von Variabilität zwischen zum Beispiel Orten, Laboratorien oder Personen. Wie in Kapitel 2.2.1 geschrieben, kann man die Messfehler in zwei Gruppen aufteilen, den klassischen und den Berksonfehler. Die richtige Einstufung eines Fehlers ist wichtig, um die Auswirkungen wie einen Bias richtig beurteilen bzw. behandeln zu können, da Methoden zur Korrektur solcher Fehler, Modelle für den Fehler selbst benötigen. Das folgende Kapitel stützt sich dabei hauptsächlich auf Heid u. a. (2004), so weit nichts anderes angegeben wird.

4.1 Definition der Fehler

Beschreibe X wieder den wahren Wert und W den gemessenen Wert. U bezeichne einen normalverteilten Messfehler mit Erwartungswert von Null, $E(U_i) = 0$, und einer homoskedastischen Varianz σ_U^2 . Dann gilt für den additiven klassischen Messfehler

$$W_i = X_i + U_i \quad , \quad (1)$$

wobei Unabhängigkeit zwischen X_i und U_i unterstellt wird. Allgemein tritt ein klassischer Messfehler auf, wenn eine Menge von einem bestimmten Gerät gemessen wird und wiederholte Messungen um den wahren Wert schwanken. Dies kann in Gleichung (1) gesehen werden, denn der gemessene Wert ist gleich dem wahren Wert plus einem Messfehler mit Erwartungswert Null.

Im Gegensatz zum klassischen Messfehler gilt für den additiven Berksonfehler

$$X_i = W_i + U_i \quad , \quad (2)$$

wobei Unabhängigkeit zwischen W_i und U_i unterstellt wird. Generell wird von einem Berksonfehler ausgegangen, wenn ein Gruppendurchschnitt jedem Individuum zugeordnet wird, zu welchem die Charakteristika der Gruppe passen. Gleichung (2) macht dies deutlich, denn dort ist der wahre Wert gleich dem (für eine Gruppe) gemessenen Wert plus einem Fehlerterm.

Ganz analog zu den beiden additiven Fällen kann jeweils auch ein multiplikativer Fehler angegeben werden. Dieser weist eine andere Struktur auf als der additive Fehler, kann aber mit einer log-Transformation leicht in einen additiven Fehler auf der log-Skala überführt werden. Es gilt

$$\begin{aligned} W_i &= X_i U_i \\ \Leftrightarrow \log(W_i) &= \log(X_i) + \log(U_i) \quad . \end{aligned}$$

Nach dieser Transformation kann dann der multiplikative Fehler wie ein additiver Fehler behandelt werden.

Heid u. a. (2004) beschreiben insgesamt drei Situationen, in welchen sie den klassischen Fehler und zwei, in welchen sie den Berksonfehler annehmen (Heid u. a., 2004, Seite 371). Erster Fall für den klassischen Fehler liegt vor, wenn wiederholte Messungen verschiedene Werte liefern und um den wahren Wert variieren. Annahme hierbei ist, dass keine anderen Fehler existieren. Die zweite Situation, in der ein klassischer Messfehler angenommen wird, tritt ein, wenn eine Messung als Ersatz bzw. Vertretung für den Durchschnitt genommen wird. Letzter Fall für den klassischen Fehler ist erfüllt, wenn Unsicherheit in der Schätzung eines Korrekturfaktors vorliegt, d.h. wenn Wiederholungen der Beobachtungen mit anderen Teilnehmern andere Korrekturfaktoren ergeben. Die eine Situation eines Berksonfehlers ist erfüllt, wenn eine Messung für eine Gruppe jedem Individuum dieser Gruppe zugeordnet wird, auch wenn sich die individuellen Werte unterscheiden. Der andere Fall eines Berksonfehlers tritt ein, wenn ein Korrekturfaktor, der für eine Gruppe von Individuen mit bestimmten Charakteristika berechnet wurde, für alle Individuen dieser Gruppe angenommen wird.

Jede Zuordnung zu einem bestimmten Fehlertyp unterliegt natürlich einer gewissen Unsicherheit. Dennoch können obige Situationen als Richtlinien gesehen werden, die eine sinnvoll erscheinende Einteilung ermöglichen. Eine genaue Klassifikation der Fehler ist notwendig, um ihre Auswirkung auf die Modelle bzw.

Ergebnisse analysieren zu können. Dies wird im folgenden Abschnitt betrachtet.

4.2 Auswirkungen der Fehler

Bei Untersuchungen der Auswirkungen von Messfehlern liegt das Hauptaugenmerk auf einem möglichen Bias. Bei annähernd jeder Studie ist der geschätzte Zusammenhang zwischen Variablen von Interesse, in epidemiologischen Untersuchungen meist der Effekt von Exposition auf die Gesundheit. Daher ist es wichtig, dass dieser Effekt korrekt geschätzt wird und keine Verzerrung aufweist. Liegt jedoch ein klassischer additiver Messfehler wie in Gleichung (1) vor, wird sich ein Bias einstellen. Für den Fall der einfachen linearen Regression gilt (Carroll u. a., 2006, Seite 43)

$$\begin{aligned}
 & Y_i = \beta_0 + \beta_1 X + v_i \\
 \Leftrightarrow & Y_i = \beta_0 + \beta_1 W_i + [\beta_1(X_i - W_i) + v_i] \\
 \Rightarrow & \hat{\beta}_1 \xrightarrow{p} \frac{\sigma_X^2}{\sigma_X^2 + \sigma_U^2} \beta_1 \quad . \quad (3)
 \end{aligned}$$

Wird der Messfehler im Modell nicht berücksichtigt, wird sich ein Bias Richtung Null ergeben, d.h. der Effekt wird kleiner geschätzt als er in Wirklichkeit ist. Dies ist in (3) ersichtlich, da der Term vor β_1 immer kleiner oder gleich 1 ist.

Anders ist der Sachverhalt, wenn ein additiver Berksonfehler vorliegt. Hier gilt (Carroll u. a., 2006, Seite 28)

$$\begin{aligned}
 E(X_i|W_i) &= W_i \\
 \text{Var}(X_i) &= \text{Var}(W_i) + \text{Var}(U_i) \quad ,
 \end{aligned}$$

und damit auch

$$\text{Var}(X_i) > \text{Var}(W_i) \quad . \quad (4)$$

Der bedingte Erwartungswert von X entspricht also dem gemessenen Wert W . Wird in den Modellen W statt X verwendet, so entsteht kein Bias in den Schätzwerten. Die Effektschätzer mit additivem Berksonfehler sind mit denen im Fall ohne Messfehler identisch. Dennoch darf auch ein additiver Berksonfehler nicht vernachlässigt werden. Denn wie in (4) ersichtlich, ist die Varianz von X größer,

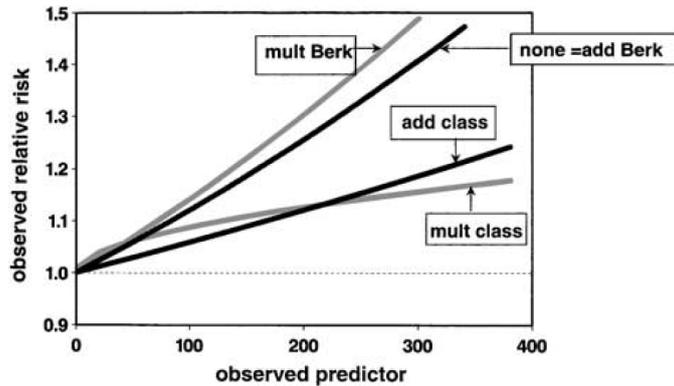


Abbildung 3: Messfehler und ihre Wirkung

die Konfidenzintervalle werden entsprechend breiter geschätzt.

Abbildung 3 (Heid u. a., 2004, Seite 374) zeigt auf recht einfache Weise, wie die verschiedenen Messfehler auf die Schätzung wirken. Als Referenz dient der Fall des additiven Berksonfehlers, da dieser keine Verzerrung aufweist. Gemäß Heid u. a. (2004) verstärkt der multiplikative Berksonfehler die Schätzung geringfügig, wenn er überhaupt einen Einfluss auf die Schätzung hat. Wie oben beschrieben wirkt der additive klassische Messfehler dämpfend auf den Effektschätzer. Im Fall des multiplikativen klassischen Messfehlers wird sogar die Krümmung der Kurve verfälscht. Daher ist es in den meisten Situationen sinnvoll den multiplikativen Messfehler durch eine log-Transformation in einen additiven umzuwandeln.

Nachdem geklärt ist, welche Typen von Messfehler existieren und welche Auswirkungen diese haben, werden nun Methoden vorgestellt, welche Messfehler im Modell berücksichtigen und behandeln können. Kapitel 5 beschäftigt sich mit dem Verfahren der Regression Calibration, welches in Kapitel 6 dann mit Maximum Likelihood und Multiple Imputation verglichen wird.

5 Regression Calibration

Im Folgenden wird das Konzept der Regression Calibration vorgestellt. Die Ausführungen beziehen sich dabei auf Carroll u. a. (2006, Kapitel 4), soweit nichts

anderes angegeben wird. Y bezeichne wieder die abhängige Variable, X den wahren Wert, W den gemessenen Wert, bzw. das Surrogat für X und Z einen Vektor mit weiteren erklärenden Variablen. Des Weiteren wird zunächst lediglich der Fall eines klassischen additiven Messfehlers betrachtet.

5.1 Algorithmus

Der Regression Calibration Algorithmus ist intuitiv gut verständlich und läuft prinzipiell in drei Schritten ab. Zuerst wird eine Regression von X auf (W, Z) durchgeführt, also der bedingte Erwartungswert $E(X|Z, W)$ gebildet. Als nächstes wird im ursprünglichen Modell für Y die Variable X durch die Schätzung von X ersetzt und die Analyse von Y dann wie gewohnt durchgeführt. Als letztes müssen die resultierenden Standardfehler mit der Bootstrap-Methode oder anderen asymptotischen Verfahren geschätzt bzw. korrigiert werden.

Regression Calibration stellt ein *working model* für die beobachteten Daten dar. Also gilt im Allgemeinen, dass die Schätzung nicht dem Erwartungswert der beobachteten Daten entspricht. Jedoch ist gerade bei linearen Modellen der Unterschied oft nur sehr gering.

5.2 Beispiel einfache lineare Regression

Um das Prinzip der Regression Calibration leichter nachvollziehbar zu gestalten, ist es sinnvoll ein Beispiel mit einer einfachen linearen Regression zu betrachten (Shifa, 2009, Seiten 44–46). Das ursprüngliche Modell für Y lautet

$$Y = \beta_0 + \beta_1 X + \epsilon \quad . \quad (5)$$

Nun wird gemäß dem Algorithmus die wahre Variable X in Abhängigkeit von der gemessenen Variable W modelliert

$$X = \alpha_0 + \alpha_1 W + u \quad . \quad (6)$$

Setzt man Gleichung (6) in Gleichung (5) ein, erhält man das Modell von Y in Abhängigkeit von W

$$Y = (\beta_0 + \alpha_0\beta_1) + \alpha_1\beta_1W + (\epsilon + \beta_1u) \quad , \quad (7)$$

oder vereinfacht

$$Y = \delta_0 + \delta_1W + v \quad . \quad (8)$$

Der interessierende Parameter ist β_1 aus Gleichung (5), da dieser den Effekt von X auf Y misst. Gleichung (7) entspricht dem naiven Modell, d.h. dem Modell wenn in Gleichung (5) fälschlicherweise W statt X verwendet wird. Man kann sehen, dass im Allgemeinen für $\alpha_1 \neq 1$ gilt

$$\delta_1 \neq \beta_1$$

und das naive Modell eine Verzerrung des Parameters verursacht. Daher ist der Ansatz der Regression Calibration von Vorteil. Denn nun können die geschätzten Parameter aus Gleichung (7) bzw. (8) durch die Parameter aus Gleichung (6) geteilt werden, da die beiden Modelle (6) und (7) im Gegensatz zu Modell (5) geschätzt und berechnet werden können. Somit ergibt sich also die Beziehung $\beta_1 = \delta_1/\alpha_1$ oder exakt

$$\hat{\beta}_{rc} = \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}}) (y_i - \bar{y})}{\sum_{i=1}^n (\hat{x}_i - \bar{\hat{x}})^2} = \frac{\sum_{i=1}^n (\hat{\alpha}_1 w_i - \hat{\alpha}_1 \bar{w}) (y_i - \bar{y})}{\sum_{i=1}^n (\hat{\alpha}_1 w_i - \hat{\alpha}_1 \bar{w})^2} = \frac{\hat{\delta}_1}{\hat{\alpha}_1} \quad . \quad (9)$$

In (9) wird β_{rc} gemäß der Standardformel für Regressionskoeffizienten als Verhältnis der Kovarianz von (X, Y) und der Varianz von X berechnet. Aus Gleichung (5) und (7) kann man erkennen, dass gilt

$$\beta_1 X = \alpha_1 \beta_1 W \quad \Rightarrow \quad X = \alpha_1 W \quad .$$

Somit können in (9) die nicht messbaren \hat{x}_i durch die messbaren $\hat{\alpha}_1 w_i$ ersetzt werden. Als nächster Schritt ist dann klar, dass sich $\hat{\beta}_{rc}$ als $\hat{\delta}_1/\hat{\alpha}_1$ ergibt und aus den beobachteten Daten geschätzt werden kann.

Abbildung 4 vergleicht die Situationen bei einem klassischen additiven Messfehler, wenn der Messfehler ignoriert oder mit dem Regression Calibration–Ansatz

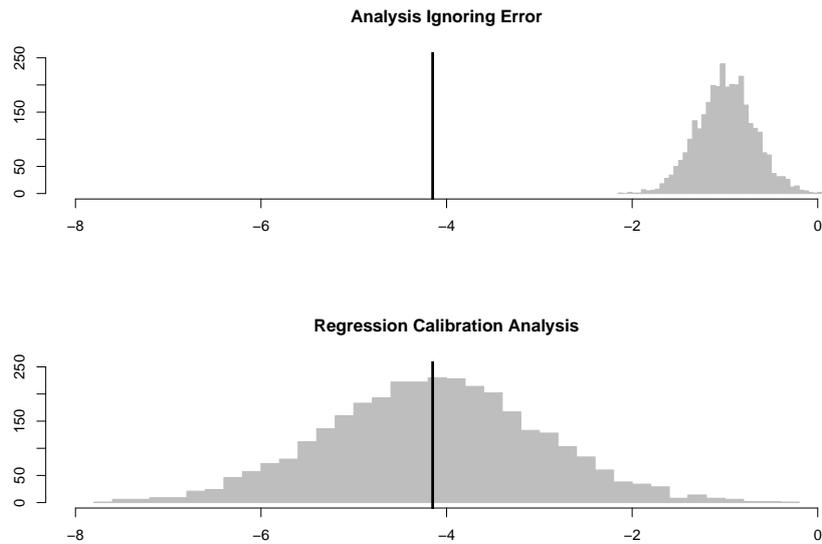


Abbildung 4: Vergleich des Bias mit und ohne Regression Calibration

behandelt wird.⁴ Man kann erkennen, dass die naive Analyse einen Bias erzeugt. Der wahre Wert liegt bei ca. -4.1, wohingegen der geschätzte Wert bei naiver Herangehensweise bei ca. -1 liegt. Dieser Bias kann durch die Modellierung über Regression Calibration vermieden werden. Jedoch besteht immer der Trade-Off zwischen einer Reduktion des Bias und einer größeren Streuung. Denn man kann sehen, dass die Parameterschätzungen beim Regression Calibration-Ansatz eine deutlich höhere Varianz aufweisen als die Schätzungen im naiven Modell.

Betrachtet man einen additiven Berkson-Messfehler, d.h. $X = W + U$, so kann man sehen, dass das naive Modell, also Gleichung (5) mit W statt X , keine Verzerrung verursacht, da wie oben beschrieben für einen Berksonfehler $E(X|W, U) = W$ gilt. Wird also X durch den bedingten Erwartungswert von X ersetzt, wird letztendlich auch mit Regression Calibration das naive Modell geschätzt. Daher entstehen in diesem Fall keine Verzerrungen.

⁴Graphik dem Beispiel von Carroll u. a. (2006, Seite 69) nachempfunden

5.3 Schätzung der Parameter

5.3.1 Drei Möglichkeiten

Generell gibt es drei verschiedene Möglichkeiten, um die Parameter der Regression Calibration zu schätzen. Der erste Ansatz kann verwendet werden, wenn *interne Validierungsdaten* vorliegen, d.h. X für einen Teil der Daten exakt gemessen werden konnte. Für den restlichen Datensatz können dann die Parameterschätzungen leicht bestimmt werden. Diese Situation wird in den folgenden Kapiteln noch genauer betrachtet. Dies ist der Idealfall, jedoch sind solche Studien gerade bei epidemiologischen Themen oftmals sehr kostenintensiv und können nicht immer im ausreichenden Umfang durchgeführt werden. Zudem kann eine derartige Situation oftmals mit dem Problem fehlender Daten gleichgesetzt werden. Hier könnten entsprechende Techniken für fehlende Daten wie Multiple Imputation einem Regression Calibration–Ansatz vorgezogen werden. Diese Problematik wird genauer in Kapitel 6 beschrieben, wenn die Methoden für Validierungsstudien miteinander verglichen werden.

Der zweite Ansatz setzt voraus, dass ein sogenanntes *unbiased instrument* vorhanden ist. Dies bedeutet, dass im Datensatz eine Variable T zur Verfügung steht, die in einem gewissen Verhältnis zu X steht. Per Definition des Begriffs gilt dann, dass die Regression von T auf (Z, W) der Regression von X auf (Z, W) entspricht. Als dritte Möglichkeit existiert der Ansatz der *Wiederholungsmessungen*. Dieser wird im Folgenden genauer dargestellt.

5.3.2 Wiederholungsmessungen

Zunächst wird nochmals kurz dargestellt wie der Regression Calibration Algorithmus abläuft. Dabei wird als erstes der bedingte Erwartungswert von X gebildet

$$E(\mathbf{X}|\mathbf{Z}, \mathbf{W}) = m_{\mathbf{X}}(\mathbf{Z}, \mathbf{W}, \gamma) \quad .$$

Die vorhergesagten Werte für X werden dann benutzt, um den bedingten Erwartungswert von Y zu bilden

$$E(\mathbf{Y}|\mathbf{Z}, \mathbf{W}) = m_{\mathbf{Y}}(\mathbf{Z}, m_{\mathbf{X}}(\mathbf{Z}, \mathbf{W}, \gamma), \mathcal{B}) \quad .$$

Hierbei bezeichnen γ Parameter, die geschätzt werden müssen und \mathcal{B} unbekannte Parameter. Allgemein gilt weiter der klassische additive Messfehler

$$W = X + U \quad \text{mit} \quad U \sim N(0, \Sigma_{uu}) \quad .$$

Für die Verteilungen von X und Z gilt (Augustin u. a., 2008, Seite 260)

$$\begin{pmatrix} X_i \\ Z_i \end{pmatrix} \stackrel{iid}{\sim} F \left(\begin{pmatrix} \mu_x \\ \mu_z \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix} \right) \quad .$$

Um letztendlich zur besten linearen Approximation zu gelangen, müssen im Vorfeld die verschiedenen Varianzmatrizen geschätzt werden. Dabei gilt, dass es k_i Messungen, also $\mathbf{W}_{i1}, \dots, \mathbf{W}_{ik}$, von \mathbf{X}_i gibt. Mit $\overline{\mathbf{W}}_{i\cdot}$ wird der jeweilige Mittelwert der k Messungen für ein \mathbf{X}_i bezeichnet.

Die Messfehlervarianz Σ_{uu} kann definiert werden, wenn sie bereits bekannt ist. Jedoch ist die Varianz in den meisten Fällen nicht bekannt und muss aus den Daten geschätzt werden. Sie ergibt sich zu

$$\hat{\Sigma}_{uu} = \frac{\sum_{i=1}^n \sum_{j=1}^{k_i} (\mathbf{W}_{ij} - \overline{\mathbf{W}}_{i\cdot}) (\mathbf{W}_{ij} - \overline{\mathbf{W}}_{i\cdot})^t}{\sum_{i=1}^n (k_i - 1)} \quad . \quad (10)$$

In Gleichung (10) ist zu erkennen, dass jeweils die quadratische Differenz zwischen dem gemessenen Wert \mathbf{W}_{ij} und dem Mittelwert über die k_i Messungen $\overline{\mathbf{W}}_{i\cdot}$ gebildet und über die Anzahl der Beobachtungen und Anzahl der Wiederholungsmessungen summiert wird. Als Korrekturfaktor ist nun nicht $n - 1$, sondern auf Grund der Wiederholungsmessungen $\sum_{i=1}^n (k_i - 1)$ zu berücksichtigen.

Die Mittelwerte für X , W und Z sind wie folgt definiert

$$\hat{\mu}_x = \hat{\mu}_w = \frac{\sum_{i=1}^n k_i \overline{\mathbf{W}}_{i\cdot}}{\sum_{i=1}^n k_i}; \quad \hat{\mu}_z = \overline{\mathbf{Z}} \quad .$$

Zudem wird ein Korrekturfaktor benötigt

$$\nu = \sum_{i=1}^n k_i - \frac{\sum_{i=1}^n k_i^2}{\sum_{i=1}^n k_i} \quad . \quad (11)$$

Nun kann die Varianz von Z und die Kovarianz von X und Z standardmäßig

bestimmt werden

$$\hat{\Sigma}_{zz} = (n-1)^{-1} \sum_{i=1}^n (\mathbf{z}_i - \bar{\mathbf{z}}) (\mathbf{z}_i - \bar{\mathbf{z}})^t, \quad (12)$$

$$\hat{\Sigma}_{xz} = \sum_{i=1}^n k_i (\bar{\mathbf{w}}_{i\cdot} - \hat{\mu}_w) (\mathbf{z}_i - \bar{\mathbf{z}})^t / \nu. \quad (13)$$

In Gleichung (12) kann man sehen, dass die Standard-Formel für die Varianzschätzung verwendet wurde. Auch in Gleichung (13) wurde die normale Formel verwendet, jedoch müssen die wiederholten Messungen mit k_i berücksichtigt werden. Zudem ist der Korrekturfaktor nicht $n-1$ sondern ν . Betrachtet man Formel (11) genauer, ist zu erkennen, dass für $k_i = 1$ sofort $\nu = n-1$ folgt, also der Standardfall eintritt, falls keine Wiederholungsmessungen vorliegen.

Für den klassischen additiven Messfehler gilt wie in Kapitel 4.2 beschrieben $E(W_i|X_i) = E(X_i)$ und $\text{Var}(W_i) = \text{Var}(X_i) + \text{Var}(U_i)$. Daher muss für die Varianz von X gelten: $\text{Var}(X_i) = \text{Var}(W_i) - \text{Var}(U_i)$. Wird dies berücksichtigt, ist die Formel für die geschätzte Varianz von X verständlich

$$\hat{\Sigma}_{xx} = \frac{1}{\nu} \left\{ \sum_{i=1}^n k_i (\bar{\mathbf{w}}_{i\cdot} - \hat{\mu}_w) (\bar{\mathbf{w}}_{i\cdot} - \hat{\mu}_w)^t \right\} - \frac{n-1}{\nu} \hat{\Sigma}_{uu}. \quad (14)$$

Gleichung (14) besteht eben aus diesen zwei Bestandteilen. Der Minuend beschreibt die geschätzte Varianz von W , wieder korrigiert um k_i und ν . Der Subtrahend repräsentiert die geschätzte Messfehlervarianz, welche ebenfalls mit einem weiteren Korrekturfaktor für die Wiederholungsmessungen versehen ist.

Mit Hilfe all dieser geschätzten Mittelwerte und Varianzen kann dann die beste lineare Approximation, bzw. die geschätzte Calibration function aufgestellt werden

$$E(\mathbf{X}_i | \mathbf{z}_i, \bar{\mathbf{w}}_{i\cdot}) \approx \hat{\mu}_w + \begin{pmatrix} \hat{\Sigma}_{xx} & \hat{\Sigma}_{xz} \\ \hat{\Sigma}_{xz}^t & \hat{\Sigma}_{zz} \end{pmatrix}^{-1} \begin{pmatrix} \bar{\mathbf{w}}_{i\cdot} - \hat{\mu}_w \\ \mathbf{z}_i - \bar{\mathbf{z}} \end{pmatrix}. \quad (15)$$

Zunächst ist zu beachten, dass es sich bei der Formel um eine Approximation handelt. Sind die Variablen gaussianisch verteilt, so beschreibt Gleichung (15) den exakten bedingten Erwartungswert. Dies ist jedoch im Allgemeinen nicht der Fall. Insgesamt besteht Gleichung (15) aus vier logischen Teilen. Der erste Sum-

mand ist der geschätzte Mittelwert von W . Der erste Teil des zweiten Summanden beschreibt die Varianz von X und die Kovarianz von X und Z . Dieser muss durch den zweiten Teil dividiert werden. Er beschreibt die Varianz-, Kovarianzstruktur von X und Z addiert mit der Messfehlervarianz, welche wieder um den Faktor k_i korrigiert werden muss. Zuletzt bezeichnet der dritte Teil die Differenz der beobachteten Werte vom Mittelwert, wobei in Variable W die Differenzen über die Mittelwerte der k verschiedenen Messungen \overline{W}_i und dem gesamten Mittelwert $\hat{\mu}_w$ gebildet werden.

Liegen also Wiederholungsmessungen vor oder kann man in der Studie wiederholte Messungen durchführen, so lassen sich die Parameter mit Gleichung (15) schätzen. Jedoch gilt dies nur unter der Annahme eines klassischen additiven Messfehlers. Im Fall eines multiplikativen Messfehlers gibt es zwei mögliche Alternativen, um eine Lösung zu erhalten. Zum Einen kann der multiplikative Fehler mit einer log-Transformation in einen additiven verwandelt werden. Zum Anderen kann die Regression Calibration aber auch angepasst werden, um sie direkt für multiplikative Fehler anwenden zu können, falls eine log-Transformation der Daten nicht erwünscht ist oder nicht sinnvoll erscheint. Dies soll hier aber nicht weiter ausgeführt werden. Ein detailliertes Vorgehen findet sich bei Carroll u. a. (2006, Seiten 72–79).

5.4 Anwendungsbereich

Regression Calibration findet in mehreren Bereichen Anwendung. Sehr oft wird sie im Rahmen der linearen Regression verwendet. Dort können oben vorgestellte Modelle benutzt werden. Aber auch die logistische Regression ist ein wichtiger Anwendungsbereich. Allerdings muss die logistische Funktion mit der Probit-Funktion approximiert werden, da die Logit-Funktion in diesem Fall keine *closed-form solution* aufweist und analytisch nicht berechnet werden kann (Carroll u. a., 2006, Seiten 90–93). Als weitere Anwendungsbereiche der Regression Calibration lassen sich die *loglinear Mean Models*, bzw. im Allgemeinen die generalisierten linearen Modelle ausmachen.

6 Vergleich Maximum Likelihood, Multiple Imputation und Regression Calibration

Die in Kapitel 5 vorgestellte Regression Calibration (RC) ist natürlich nicht die einzige Möglichkeit, um Messfehler in Modellen zu behandeln. Genauso kann die Methode der Maximum Likelihood (ML) oder der Multiple Imputation (MI) benutzt werden. Im Folgenden sollen alle drei Konzepte miteinander verglichen werden. Die Ausführungen beziehen sich dabei auf Messer u. Natarajan (2008), soweit nichts anderes angegeben wird.

6.1 Das Modell

Zur Analyse wird ein Modell der logistischen Regression betrachtet mit einer binären abhängigen (Krankheits-)Variable Y . Diese hängt von der Exposition X und weiteren Variablen Z ab. Das Modell für die logistische Regression lautet standardmäßig

$$P(Y = 1|X, Z) = \pi(X\beta_x + Z\beta_z) \equiv \frac{\exp(X\beta_x + Z\beta_z)}{1 + \exp(X\beta_x + Z\beta_z)} . \quad (16)$$

Der Parameter von Interesse ist β_x , da dieser das Risiko der Krankheit bei einer gegebenen Exposition quantifiziert. Wie in den vorigen Kapiteln beschrieben kann die Exposition X nicht immer beobachtet oder nur mit Fehler gemessen werden. Daher wird das Surrogat W betrachtet. Für den Zusammenhang mit X wird eine multiple lineare Regression angenommen

$$W = X\alpha_x + Z\alpha_z + \epsilon_w \quad \text{mit} \quad \epsilon_w \sim N(0, \sigma_w^2 I) . \quad (17)$$

Dabei bezeichnet ϵ_w den Messfehler und α die jeweiligen Matrizen der Regressionskoeffizienten.

Zumeist wird das Surrogat W dazu benutzt, um den unbeobachteten Wert der Exposition X selbst vorherzusagen. Daher ist die bedingte Verteilung von X gegeben W von Interesse. Unter der Annahme einer multivariaten Normalverteilung für X gilt, dass ebenso die bedingte Verteilung von X gegeben W einer multiva-

riaten Normalverteilung folgt und somit für X gilt

$$X = W\gamma_w + Z\gamma_z + \epsilon_x \quad \text{mit} \quad \epsilon_x \sim N(0, \Sigma_{x|w}) \quad . \quad (18)$$

Die Parameter γ und $\Sigma_{x|w}$ können dabei auch durch die Parameter α , σ_w^2 und Σ_s ausgedrückt werden. Man erhält die Informationen dieser Messfehlerparameter aus der sogenannten Validierungsstudie. Eine Validierungsstudie ist eine Studie mit geringerem Umfang als die Hauptstudie, bei welcher die Exposition X und das Surrogat W beobachtet werden. Dabei unterscheidet man zwischen *interner* und *externer* Validierungsstudie. Erstere besteht aus einer Teilstichprobe der Hauptstudie mit den selben Personen und es werden die Variablen Y , X und W gemessen. In der Hauptstudie werden hingegen immer nur Y und W beobachtet. Bei einer externen Validierungsstudie benutzen die Validierungs- und die Hauptstudie unabhängige Stichproben mit zum Teil unterschiedlichen Teilnehmern. Dabei wird in der Validierungsstudie X und W gemessen, in der Hauptstudie Y und W .

Im Folgenden werden die weiteren Variablen Z in der Notation nicht berücksichtigt, auch wenn sie eigentlich im Modell vorhanden sind. Dies dient der Übersichtlichkeit für die benutzten Formeln und Modelle.

6.2 Maximum Likelihood

Maximum Likelihood wird oft als Standard-herangehensweise für Messfehlerprobleme gesehen. Theoretisch betrachtet ist sie die genaueste Methode und daher gegenüber anderen Verfahren zu bevorzugen. Jedoch können starke Verzerrungen auftreten, wenn die Verteilungsannahme für die Variablen und bedingten Verteilungen falsch ist oder stark von der wahren Verteilung abweicht.

6.2.1 Likelihood für X beobachtet

Für den Fall, dass X beobachtet werden kann, ist Y unabhängig von dem Surrogat W , es gilt also $P(Y|X, W) = P(Y|X)$. Weiter gilt demnach $P(Y, X|W) = P(Y|X, W)P(X|W) = P(Y|X)P(X|W)$. Wird dies beachtet, ergibt sich die log-Likelihood für die gesamten Daten zu

$$l(\beta, \gamma \Sigma_{x|w}; Y, X|W) = l_1(\beta; Y|X) + l_2(\gamma, \Sigma_{x|w}; X|W) \quad . \quad (19)$$

Hierbei beschreibt l_1 die normale Likelihood einer logistischen Regression, wenn $N = 1$ gilt, also

$$l_1(\beta; Y|X) = YX\beta - \log(1 + \exp(X\beta)) \quad . \quad (20)$$

Analog bezeichnet l_2 die Likelihood einer der Normalverteilung folgenden Regression mit $N = 1$, also

$$l_2(\gamma, \Sigma_{x|w}; X|W) = -\frac{1}{2} \left(\log(\det(\Sigma_{x|w})) + (X - W\gamma_w)^t \Sigma_{x|w}^{-1} (X - W\gamma_w) \right) \quad . \quad (21)$$

6.2.2 Likelihood für X nicht beobachtet

Die Likelihood aus Kapitel 6.2.1 geht davon aus, dass die Exposition X beobachtet werden kann. Da aber von einem Design mit Haupt- und Validierungsstudie ausgegangen wird, ist X für den Großteil der Daten nicht beobachtbar. Daher muss auch die Likelihood noch entsprechend erweitert werden.

Zunächst einmal kann die marginale Verteilung $P(Y|W)$ bestimmt werden. Diese lässt sich berechnen, indem in der gemeinsamen Wahrscheinlichkeitsverteilung $P(Y, X|W)$ über die unbeobachteten X integriert wird

$$P(Y|W; \beta, \gamma, \Sigma_{x|w}) = \int P(Y|x; \beta)P(x|W; \gamma, \Sigma_{x|w})dx \quad . \quad (22)$$

Gleichung (22) ist zugleich aber auch die Likelihood des Modells, welches sich ergibt, wenn Gleichung (18) in (16) eingesetzt wird

$$P(Y = 1|W, \epsilon_x; \beta, \gamma) = \pi((W\gamma)\beta + \epsilon_x\beta) \quad ,$$

da X gemäß Gleichung (18) von W , γ und ϵ_x abhängt. Zur Vereinfachung wird im Folgenden $u = \epsilon_x\beta$ gesetzt. Somit kann die log-Likelihood geschrieben werden als

$$\begin{aligned}
l_3(\beta, \gamma, \Sigma_{x|w}; Y|W) &= \\
&= \log \left((\beta^t \Sigma_{x|w} \beta)^{-\frac{1}{2}} \int \frac{\exp((W\gamma\beta + u)Y)}{1 + \exp(W\gamma\beta + u)} \exp\left(-\frac{1}{2}u^t(\beta^t \Sigma_{x|w} \beta)^{-1}u\right) du \right) .
\end{aligned} \tag{23}$$

Gleichung (23) beschreibt also die Likelihood für den Fall, dass X nicht beobachtbar ist.

6.2.3 Likelihood für gesamte Stichprobe

Für die gesamte Stichprobe ergibt sich die log-Likelihood als Summe über die beiden log-Likelihoods (19) und (23)

$$\begin{aligned}
l(\beta, \gamma, \Sigma_{x|w}) &= \\
&= \sum_{i \in \text{validation}} (l_1(\beta; Y_i|X_i) + l_2(\gamma, \Sigma_{x|w}; X_i|W_i)) + \sum_{i \in \text{main}} l_3(\beta, \gamma, \Sigma_{x|w}; Y_i|W_i) .
\end{aligned} \tag{24}$$

Im linken Summanden wird die Likelihood aus der internen Validierungsstudie zusammengefasst, wenn sowohl Y als auch X und W beobachtet werden können. Hinzu kommt die Likelihood des rechten Summanden, welche die Likelihood der Hauptstudie widerspiegelt. Dort ist X nicht beobachtbar. Liegt eine externe Validierungsstudie vor, fällt der Term l_1 weg, da Y dort nicht beobachtet wird und der Term l_1 bei der Integration über das fehlende Y gleich Null wird.

Unter der Annahme, dass das Modell (24) korrekt ist, sind die Maximum Likelihood Schätzer konsistent, annähernd normalverteilt und weisen den kleinsten asymptotischen MSE auf. Daher setzen diese ML-Schätzer für β den *Goldstandard* für alle zum Vergleich herangezogenen Methoden.

Das Problem der ML-Methode liegt jedoch darin, dass sie nicht in geschlossener Form berechenbar ist. Daher müssen numerische Optimierungsverfahren herangezogen werden wie zum Beispiel der Newton-Raphson-Algorithmus. Hinzu kommt, dass das Integral in (23) jedes mal numerisch approximiert werden muss, wenn die Likelihood berechnet wird. All dies kann bei größeren Stichproben zu längeren Rechenzeiten führen, weswegen im Folgenden noch zwei andere Methoden vorgestellt werden.

6.3 Regression Calibration

In Kapitel 5 wurde die Regression Calibration bereits vorgestellt. Das Hauptaugenmerk wurde dabei auf die Schätzung der Parameter bei Wiederholungsmessungen gelegt. Nun liegen aber Validierungsdaten vor und es soll die Likelihood betrachtet werden. Das generelle Vorgehen der Regression Calibration bleibt jedoch das selbe.

Zunächst einmal muss mit Hilfe der Validierungsdaten die Likelihood des Messmodells (18) maximiert werden, um die Regressionsschätzer $\hat{\gamma}$ und $\hat{\Sigma}_{x|w}$ zu erhalten. Dieser Schritt ist für interne wie für externe Validierungsdaten identisch

$$\sum_{i \in \text{validation}} l_2(\gamma, \Sigma_{x|w}; X_i | W_i) \quad .$$

Im zweiten Schritt werden diese geschätzten Parameter herangezogen, um den bedingten Erwartungswert $E(X|W)$ zu bestimmen. Für alle Beobachtungen, bei welchen X nicht beobachtet wurde, wird dann der geschätzte bedingte Erwartungswert von X , $\hat{E}(X|W)$ eingesetzt. Somit ergibt sich für die approximierte Likelihood

$$l_{RC}(\beta) = \sum_{i \in \text{validation}} l_1(\beta; Y_i | X_i) + \sum_{i \in \text{main}} l_1(\beta; Y_i | \hat{E}(X | W_i)) \quad . \quad (25)$$

Man kann erkennen, dass die Likelihood (25) aus zwei Teilen besteht. Der linke Teil beschreibt all die Beobachtungen aus der Validierungsstudie, für welche X gemessen werden konnte. Hinzu kommt der rechte Teil, welcher alle Beobachtungen der Hauptstudie einschließt, für welche X nicht beobachtet wurde und daher X mit dem geschätzten bedingten Erwartungswert von X ersetzt werden musste. Des Weiteren sind die Likelihoods l_1 und l_2 wie in Gleichung (20) bzw. (21) definiert.

Vergleicht man die Likelihood der Regression Calibration (25) mit der Likelihood der Maximum-Likelihood-Methode (24), fallen zwei Problematiken auf. Zum Einen werden bei der Regression Calibration die Informationen der Messfehler-Parameter γ und Σ ignoriert, die in der gemischten Verteilung l_3 enthalten sind. Zum Anderen wird der Term l_3 durch die an der Stelle $\hat{E}(X|W)$ ausgewertete logistische Likelihood l_1 approximiert. Daher ist der Regression-Calibration-Ansatz rechentechnisch einfacher als Maximum Likelihood. Jedoch führt dies dazu, dass

die RC-Schätzer asymptotisch weniger effizient sind als die entsprechenden ML-Schätzer.

Messer u. Natarajan (2008, Seiten 5–7) zeigen, dass die Schätzer der Regression Calibration auf Grund der ersten Problematik nur dann an Effizienz verlieren, wenn der Term l_3 tatsächlich zusätzliche Informationen über γ und Σ enthält. Dies ist jedoch lediglich dann der Fall, wenn $\beta' \Sigma \beta$ groß ist und nur in dieser Situation sind Effizienzverluste der RC-Schätzer zu erwarten. Ähnlich verhält es sich mit der zweiten Problematik. Ist der Term $\beta^t \Sigma_{x|w} \beta$ klein, liegt die RC-Likelihood (25) sehr nahe an der Likelihood (24). Die Schätzer werden in diesem Fall numerisch annähernd identisch sein. Dies kommt daher, dass der limitierende Faktor in $l_3(Y|W)$ in der Tat der Wert $l_1(Y|E(X|W))$ ist für den Fall, dass $\beta^t \Sigma_{x|w} \beta \rightarrow 0$ gilt. Ist $\beta^t \Sigma_{x|w} \beta$ hingegen groß, kann es passieren, dass die RC-Schätzer die falsche Gleichung lösen und es ist mit größeren Abweichungen der Ergebnisse zu rechnen.

6.4 Multiple Imputation

Multiple Imputation ist ein Verfahren, um fehlende Daten in Stichproben zu behandeln. Wie der Name sagt, werden für jeden fehlenden Wert in m Iterationen Werte geschätzt, welche die fehlenden wahren Werte bei der Analyse ersetzen. Der Mittelwert der sich ergebenden m Parameterschätzer stellt dann den finalen Schätzer dar. In einem Studiendesign mit Haupt- und Validierungsstudie kann Multiple Imputation also auch zur Behandlung von Messfehlern benutzt werden, da die Werte für X nur in der Validierungsstudie gemessen wurden und in der Hauptstudie fehlen. Somit tritt ein Problem von fehlenden Daten auf. Wichtig dabei ist, dass die Daten zufällig fehlen (*missing at random*) und kein Muster aufweisen. Dies ist bei Validierungsstudien der Fall, wenn die Personen zufällig in die Haupt- oder Validierungsstudie gezogen werden. Weiter gilt, dass die Verteilung $X|Y, W, Z$ immer die gleiche ist, egal ob X gemessen werden konnte oder nicht. Für den Fall einer externen Validierungsstudie kann diese bedingte Verteilung jedoch nicht beobachtet werden und die Behauptung der zufällig fehlenden Werte hängt mehr von a priori Modellannahmen ab.

Sowohl die zugeschriebenen Daten (*imputed data*) als auch die Parameterwerte werden mit Hilfe von Markov-Chain-Monte-Carlo-Verfahren (MCMC) bestimmt. Das Ergebnis der Markov-Kette bei jeder Iteration besteht dabei aus

einem zufällig gezogenen Parametervektor (Parameter-Schritt) und einem Vektor der zugeschriebenen Daten (Imputations-Schritt). Beim Iterationsschritt $(t - 1)$ beschreiben $(\beta, \gamma, \Sigma_{x|w})^{(t-1)}$ die aktuellen Parameter und $\tilde{X}(Y_i, W_i)^{(t-1)}$ die entsprechenden zugeschriebenen Daten für die i -te Beobachtung. Letztere hängen dabei von den beobachteten Werten Y_i und W_i ab. Im nächsten Schritt berechnet die Markov-Kette die Parameterschätzer $(\hat{\beta}, \hat{\gamma}, \hat{\Sigma}_{x|w})^{(t)}$, indem die Likelihood der vervollständigten Daten maximiert wird. Diese Likelihood ergibt sich bei internen Validierungsdaten zu

$$l_{MI} = \sum_{i \in \text{validation}} l_1(\beta; Y_i, X_i) + l_2(\gamma, \Sigma_{x|w}; X_i, W_i) + \sum_{i \in \text{main}} l_1(\beta; Y_i, \tilde{X}(Y_i, W_i)^{(t-1)}) + l_2(\gamma, \Sigma_{x|w}; \tilde{X}(Y_i, W_i)^{(t-1)}, W_i) \quad . \quad (26)$$

Dabei sind l_1 und l_2 wieder wie in Gleichung (20) bzw. (21) definiert. Es fällt auf, dass die Summe über die Validierungsdaten in Gleichung (26) mit der entsprechenden Summe in Gleichung (24) identisch ist. Zudem ist in (26) die Summe über die Hauptstudien Daten gleich der Summe über die Validierungsdaten mit der Ausnahme, dass statt X die zugeschriebenen Daten $\tilde{X}(Y_i, W_i)^{(t-1)}$ verwendet werden.

Für den Imputationsschritt einer Hauptstudienbeobachtung wird dann die bedingte Verteilung der fehlenden X -Werte gegeben Y_i und W_i benötigt, ausgewertet an der Stelle $(\beta, \gamma, \Sigma_{x|w})^{(t)}$. Die Markov-Kette berechnet $\tilde{X}(Y_i, W_i)^{(t)}$ als zufällige Ziehung aus dieser bedingten Verteilung, welche proportional ist zu

$$f(x) = \frac{\exp(x\beta^{(t)}Y_i)}{1 + \exp(x\beta^{(t)})} \exp\left(-\frac{1}{2} (x - W_i\gamma^{(t)})' \left(\Sigma_{x|w}^{(t)}\right)^{-1} (x - W_i\gamma^{(t)})\right) \quad . \quad (27)$$

Diese Verteilung ist jedoch standardmäßig in keiner Software verfügbar. Daher muss sie zumeist approximiert werden. Für genauere Details sei auf Messer u. Natarajan (2008, Seiten 8–9) verwiesen.

Das Verfahren der Multiple Imputation mit Hilfe des MCMC-Verfahrens benutzt die Tatsache, dass die Ergebnisse der Markovkette, d.h. $(\beta, \gamma, \Sigma_{x|w}, \tilde{X})^{(t)}$ mit $t = b + k, b + 2k, \dots$, approximativ unabhängig identisch verteilte Ziehungen einer limitierenden Verteilung sind. Dabei beschreibt b die Anzahl der Iterationen

der Markov-Kette und k die Anzahl der Iterationen der Kette zwischen den Imputationen, wobei beide Parameter relativ groß gewählt werden. Die limitierende Verteilung für $\tilde{X}^{(t)}$ ist eben genau das Wahrscheinlichkeitsmodell aus Gleichung (27), ausgewertet an den ML-Schätzern.

Um dann die MI-Schätzer zu erhalten, werden m mit zugeschriebenen Daten vervollständigte Datensätze abgespeichert, nachdem ausreichend viele Iterationen durchgeführt wurden. Mit Hilfe von Gleichung (26) kann dann für jeden dieser vervollständigten Datensätze der Maximum-Likelihood-Schätzer $\hat{\beta}^{(i)}$ berechnet werden. Diese können mit den gleichen Methoden ermittelt werden, die benutzt würden, wenn alle Daten von Anfang an vorhanden gewesen wären. Der eigentliche MI-Schätzer ergibt sich dann als Durchschnitt all dieser Imputations-Schätzer als

$$\hat{\beta}_{MI} = \frac{1}{m} \sum_{i=1}^m \hat{\beta}^{(i)} \quad .$$

Auf den ersten Blick verwunderlich scheint dabei vielleicht, dass die Anzahl der vervollständigten Datensätze m und damit die Anzahl der Imputations-Schätzer meistens lediglich in dem Bereich zwischen 3–10 liegt. Dies ist auf das Rao-Blackwell-Theorem zurückzuführen (Blackwell, 1947). Es muss beachtet werden, dass die Imputations-Variabilität der $\hat{\beta}^i$ klein ist, da nur der Teil von Iteration zu Iteration variiert, welcher in den Daten auch tatsächlich fehlt. Formal kann man schreiben

$$E(\beta^{(i)}) = E\left(E\left(\beta^{(i)} | \tilde{X}^{(i-1)}\right)\right) \quad ,$$

wobei das Rao-Blackwell-Theorem sicherstellt, dass gilt

$$\text{Var}\left(E\left(\beta^{(i)} | \tilde{X}^{(i-1)}\right)\right) < \text{Var}(\beta^{(i)}) \quad .$$

Für weiterführende Erklärungen zu dieser Thematik bei Multiple Imputation sei auf Schafer (1997) verwiesen.

Für externe Validierungsstudien kann der Ansatz so wie beschrieben nicht einfach verwendet werden, da Y in der Validierungsstudie nicht beobachtet wurde. Daher könnte man sich als Erweiterung für externe Validierungsdaten vorstellen auch die fehlenden Y -Werte mit Multiple Imputation zu schätzen. Die Likelihood wäre dann die selbe wie in Gleichung (26) mit der Ausnahme, dass die Y -Werte in

der Summe über die Validierungsstudie durch die geschätzten $\tilde{Y}(X_i, W_i)$ ersetzt werden müssten.

6.5 Simulation und Vergleich

Nachdem die drei Methoden Maximum Likelihood, Regression Calibration und Multiple Imputation vorgestellt wurden, werden sie in einem nächsten Schritt miteinander verglichen. Hierzu wird in Kapitel 6.5.1 kurz erläutert, welche Daten benutzt bzw. wie sie simuliert wurden, um die Ergebnisse der drei Methoden dann in Kapitel 6.5.2 zu analysieren und einander gegenüber zu stellen.

6.5.1 Vorgehen

Die Simulationen legen alle das vorgestellte Modell (16) zu Grunde. Dabei werden zwei Expositionsvariablen X_1 und X_2 simuliert. Hinzu kommt eine Variable Z , die ohne Fehler für alle Beobachtungen gemessen werden kann. Insgesamt wurden 1000 Datensätze mit je 500 Beobachtungen simuliert gemäß

$$Y|(X_1, X_2, Z) \sim \text{Bern}(\eta) \quad \text{mit} \quad \text{logit}(\eta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 Z \quad ,$$

wobei gilt $(X_1, X_2, Z) \sim N(0, \Sigma)$. Hinzu kommen dann die Surrogate W_1 bzw. W_2 für die Expositionsvariablen, welche entsprechend einem Fehlermodell wie in Gleichung (17) simuliert wurden

$$\begin{aligned} W_1|X_1 &\sim N(\alpha_{01} + \alpha_{11}X_1, \sigma_{w|x}^2) \quad , \\ W_2|X_2 &\sim N(\alpha_{02} + \alpha_{12}X_2, \sigma_{w|x}^2) \quad . \end{aligned} \quad (28)$$

Da die Methoden für die Situationen mit Validierungsdaten verglichen werden sollten, wurden jeweils für die Hälfte der Beobachtungen die Werte für X_1 und X_2 gelöscht. Dieser Fall entspricht also internen Validierungsdaten, da Y , X , W und Z für die Validierungsstudie vorliegen und lediglich Y , W und Z für die Hauptstudie. Um den Fall einer externen Validierungsstudie zu simulieren, wurden dann für die 250 Beobachtungen der Validierungsstudie zusätzlich die Y -Werte gelöscht. Auf diesem Wege können beide Arten von Validierungsstudien erzeugt werden.

Natürlich sollen auch verschiedene Szenarien betrachtet werden, um den Ver-

gleich für möglichst viele Situationen durchführen zu können. Dabei werden verschiedene Annahmen an die Parameter getroffen. Zunächst einmal gilt für alle fünf folgenden Modelle

$$\alpha_{01} = \alpha_{02} = 0 \quad \text{und} \quad \alpha_{11} = \alpha_{12} = 1 \quad .$$

Somit ist sichergestellt, dass die Variablen W_1 bzw. W_2 unverzerrte Surrogate für X_1 bzw. X_2 sind, da mit diesen Annahmen die Mittelwerte der Normalverteilung aus (28) auf X_1 bzw. X_2 reduziert werden. Des Weiteren werden die unabhängigen Variablen X_1 , X_2 und Z so standardisiert, dass sie Mittelwert Null und Varianz Eins besitzen. Zudem sind X_2 und Z unabhängig, so dass sich stets ein Korrelationskoeffizient $\rho_{2z} = 0$ ergibt. Weiter wird in dem Modell $\beta_0 = -1$ gesetzt.

In Modellen 1–4 gilt $\beta_1 = 1$, in Modell 5 $\beta_1 = 3$. Variiert wird außerdem bei den Standardabweichungen der Messfehler $\sigma_{w|x}$. Modell 1, 2 und 5 besitzen einen relativ hohen Messfehler mit $\sigma_{w|x} = 3$, Modell 3 und 4 einen geringeren mit $\sigma_{w|x} = 1$. Weitere Unterschiede werden über die Korrelationen zwischen X_1 und X_2 bzw. zwischen X_1 und Z erzeugt. Modell 1 und 3 werden mit einer Korrelation von $\rho_{12} = 0.2$ bzw. $\rho_{1z} = 0.7$ gerechnet, Modell 2, 4 und 5 umgekehrt mit Korrelationen von $\rho_{12} = 0.7$ und $\rho_{1z} = 0.2$. Tabelle 4 schafft einen kleinen Überblick über die verschiedenen Modelle und deren Parametereigenschaften.

Modell	β_1	$\sigma_{w x}$	ρ_{12}	ρ_{1z}
1	1	3	0.2	0.7
2	1	3	0.7	0.2
3	1	1	0.2	0.7
4	1	1	0.7	0.2
5	3	3	0.7	0.2

Tabelle 4: Übersicht über die fünf verschiedenen Modelle und deren Parameter

Modell 5 gilt dabei als *worst case scenario*, da hier die extremsten Werte angenommen wurden. Generell aber betonen Messer u. Natarajan (2008), dass die gewählten Parameter für epidemiologische Studien realistisch seien und auch Modell 5 nicht komplett unrealistisch erscheint. Zudem sei es nützlich die verschiedenen Methoden (ML, RC und MI) auch in einem worst case scenario zu untersuchen, um feststellen zu können welche Methode unter den extremen Bedingungen die

besten Ergebnisse liefert. Denn in Realität ist nie völlig klar, ob nicht ein solcher Extremfall vorliegt oder nicht.

Der Vorteil von Simulationsstudien wie dieser ist es, dass die wahren Werte ebenfalls vorhanden sind. Daher kann die Auswertung der Datensätze mit kompletten Beobachtungen und Variablen als *Goldstandard* herangezogen werden. Alle untersuchten Methoden, die zur Behandlung von Messfehlern oder fehlenden Daten verwendet werden, müssen sich an diesem messen lassen.

6.5.2 Ergebnisse

Nachdem klar ist, wie die Daten erzeugt wurden und welche verschiedenen Szenarien untersucht werden, kann die Analyse und der Vergleich der drei Methoden durchgeführt werden. Um die drei Methoden direkt miteinander in Verbindung zu setzen, wird ein relativer RMSE (root mean squared error) gebildet. Dieser beschreibt das Verhältnis zwischen dem RMSE der Regression Calibration bzw. der Multiple Imputation und dem RMSE der Maximum Likelihood. Somit ist ohne Umwege ein Vergleich der Methoden möglich. Dabei beschreiben Werte kleiner als Eins Situationen, in welchen die entsprechende Methode bessere Ergebnisse liefert als die Maximum-Likelihood-Methode.

In Tabelle 5 werden die Ergebnisse für die primäre Expositionsvariable β_1 dargestellt. Für die Multiple Imputation sind nur Ergebnisse für den Fall einer internen Validierungsstudie angegeben, da wie oben beschrieben für externe Designs die Methode erst genauer angepasst werden müsste. Die exakten numerischen Ergebnisse der Methoden mit Vergleich zu dem Goldstandard sind in Tabelle 9 für interne Validierungsstudien und in Tabelle 10 für externe Validierungsstudien in Anhang I zu finden. Für einen Vergleich der Methoden untereinander ist jedoch Tabelle 5 ausreichend und auch übersichtlicher.

Zunächst fällt auf, dass für Modelle 1–4 der MI-Ansatz nicht besser und nicht schlechter als der ML-Ansatz scheint. In Tabelle 9 kann man ebenfalls sehen, dass die Ergebnisse auch numerisch sehr ähnlich sind, was daraufhin deutet, dass die Approximation der bedingten Verteilung für Multiple Imputation bei internen Designs sehr gut ist. Alleine im extremen Fall von Modell 5 zeigen die ML-Schätzer einen ca. 15 prozentigen Vorteil gegenüber den MI-Schätzern.

Für die Modelle 1 und 2 mit größerem Messfehler und internem Design ist auch

Modell	$\beta' \Sigma_{x w} \beta$	MI (intern)	RC (intern)	RC (extern)
1	0.73	0.99	1.01	1.03
2	1.37	0.99	1.01	1.48
3	0.45	1.02	1.10	0.72
4	0.63	1.00	1.11	0.71
5	8.84	1.15	1.28	3.89

Tabelle 5: Relativer RMSE von RC und MI zu ML für β_1

der Regression Calibration Ansatz etwa gleich gut wie die Methode der Maximum Likelihood. Etwas schlechter scheinen die RC-Schätzer im Fall mit kleineren Messfehlern zu funktionieren. Hier kann bei den ML-Schätzern in Modell 3 und 4 ein ca. 10 prozentiger Vorteil ausgemacht werden. Dieser ist hauptsächlich auf eine größere Standardabweichung der RC-Schätzer zurückzuführen, da der Bias der beiden Methoden in etwa die gleiche Größenordnung besitzt (siehe fett markierte Werte in Tabelle 9). Auch in Modell 5 sind die Schätzer der Regression Calibration wesentlich weniger effizient als die der Maximum Likelihood, was zu einem erheblich schlechteren Ergebnis für diesen Extremfall führt.

Andere Ergebnisse ergeben sich für die Situation einer externen Validierungsstudie. In einigen Fällen lagen hier große Ausreißer vor und zum Teil sehr schiefe Verteilungen. Erwartungsgemäß waren der Bias für Modelle 1 und 2 auf Grund der großen Messfehler sowohl bei RC als auch bei ML wesentlich stärker als in Modellen 3 und 4 (siehe Tabelle 10). Für Modell 1 erzielten die RC-Schätzer und die ML-Schätzer in etwa die gleiche Standardabweichung und RMSE, weswegen die Methoden in diesem Fall als gleichwertig einzuschätzen sind. Dies ist auf die hohe Korrelation zwischen X_1 und Z zurückzuführen und somit auf einen eher geringen Fehler bei der Prognose von X gegeben W und Z bei der Methode der Regression Calibration.

Wie in Kapitel 6.3 beschrieben wurde, ist davon auszugehen, dass Regression Calibration am besten funktioniert, wenn der Term $\beta' \Sigma_{x|w} \beta$ klein ist. Betrachtet man die Modelle 3 und 4, welche einen geringen Messfehler aufweisen, kann man sehen, dass die RC-Schätzer wesentlich bessere Ergebnisse liefern als die ML-Schätzer. Hier nimmt der beschriebene Term $\beta' \Sigma_{x|w} \beta$ die kleinsten Werte in den Simulationen an. In diesen beiden Modellen weisen die RC-Schätzer nicht nur einen kleineren Bias auf, sondern zudem eine geringere Standardabweichung als

die ML-Schätzer (siehe fett markierte Werte in Tabelle 10). Dies führt zu der deutlich besseren Effizienz des RC-Ansatzes. Messer u. Natarajan (2008) konnten bei genauerer Betrachtung feststellen, dass die ML-Schätzer eine größere Schiefe und mehr Ausreißer aufwiesen. Dadurch wurden die Verteilungsannahmen wie oben beschrieben stark verletzt und die ML-Schätzung verursacht eine stärkere Verzerrung. Wurden die 10 Prozent extremsten Schätzer entfernt, näherte sich das Ergebnis der ML-Schätzungen sehr stark an das der RC-Schätzungen an. Weiter konnten die Autoren beobachten, dass bei einer Verdopplung des Stichprobenumfangs auf 1000 in der Hauptstudie und 500 in der Validierungsstudie die ML-Schätzer die RC-Schätzer an Effizienz überholen konnten.

Für Modelle 2 und 5 nimmt der Term $\beta' \Sigma_{x|w} \beta$ verhältnismäßig große Werte an. Daher ist zu erwarten, dass der RC-Ansatz in diesen Situationen eher schlechtere Ergebnisse liefert. Wie man in Tabelle 5 erkennen kann, ist genau dies der Fall. Liegen große Messfehler vor und ist X mit Z nicht stark korreliert, funktioniert die Methode der Regression Calibration nicht sehr gut. Hier ist die ML-Methode vorzuziehen. Vor allem im worst-case-Szenario von Modell 5 sind die Ergebnisse der Regression Calibration auf Grund von sehr hohen Standardabweichungen schwach.

Zusammenfassend ist zu sagen, dass in Situationen einer internen Validierungsstudie Multiple Imputation genauso wie Maximum Likelihood verwendet werden kann, solange keine extremen Szenarien vorliegen. Ebenso liefert Regression Calibration gute Ergebnisse, auch wenn die Methode der Maximum Likelihood einen gewissen Vorteil aufweist. Für externe Validierungsstudien kann sowohl Regression Calibration als auch Maximum Likelihood besser geeignet sein. Hier kommt es vor allem auf die Größe des Messfehlers, die Korrelation zwischen Y , X und Z , sowie auf den Stichprobenumfang an. Ist letzterer groß genug, so scheinen die Vorteile stets bei dem ML-Ansatz zu liegen. Interessant ist noch zu erwähnen, dass die relativen Vorteile der einzelnen Methoden hauptsächlich aus unterschiedlich großen Standardabweichungen entstanden und nicht auf Grund von verschiedenen Biases. Der größte Nachteil der Studie von Messer u. Natarajan (2008) liegt darin, dass der Umfang der Validierungsstudie 50 % des Hauptstudienumfangs beträgt. Bei epidemiologischen Studien ist dieser Anteil meist deutlich kleiner. Ein Ansatz für weitere Forschung wäre daher zu testen, ob sich die Ergebnisse verändern, wenn der Umfang der Validierungsstudie verhältnismäßig kleiner ausfällt.

7 Regression Calibration in der Praxis

Im vorangegangenen Kapitel wurde gezeigt, dass der Regression Calibration Ansatz oftmals nicht wesentlich schlechtere Ergebnisse liefert als die Methode der Maximum Likelihood und teilweise sogar besser ist. Da Regression Calibration intuitiver, einfacher und schneller anzuwenden ist als Maximum Likelihood, kommt es in der Praxis immer wieder zum Einsatz. Wie Carroll u. a. (2006, Seite 65) schreiben, ist die Einfachheit von Regression Calibration ihr größter Vorteil und die Ergebnisse sind oft überraschend gut. Im folgenden Abschnitt werden zwei Studien vorgestellt, welche zur Analyse die Methode der Regression Calibration herangezogen haben. Diese sind zum Einen Strand u. a. (2007) und zum Anderen Spiegelman (2010).

7.1 Studie von Strand et al.

In der Studie von Strand u. a. (2007) wurden die Gesundheitseffekte von Feinstaubbelastung auf den menschlichen Körper untersucht. Als *health outcome*, also als die Variable, die den Gesundheitszustand eines Menschen beschreibt, wurde die Ein-Sekunden-Kapazität (Die Ein-Sekunden-Kapazität ist ein Messparameter für die Lungenfunktionalität, welche maßgeblich durch Feinstaubbelastung beeinflusst wird). gewählt. Die Feinstaubbelastung wird anhand des $PM_{2.5}$ gemessen. Die Messung des $PM_{2.5}$ erfolgt weitestgehend über Außenmessstellen. Letztendlich ist zwar die persönliche Belastung von Interesse, doch müssen immer Rückschlüsse auf die Werte an den Außenmessstellen gezogen werden, da nur diese über Qualitätsstandards reguliert werden können.

Das Ziel der Studie ist es verschiedene Methoden zur Schätzung des health outcomes zu evaluieren. Dabei wird die Methode der *positive matrix factorization* (PMF) und der *Extrapolation* betrachtet. Bei letzterer werden nochmals verschiedene Parameter und Feinstaubbestandteile untersucht. Hierbei können Probleme auftreten, da es viele Außenmessstellen gibt, aber kaum Messungen für die Belastung im Haus oder der persönlichen Belastung. Zudem kann es schwer fallen die Quellen des Feinstaubes zu unterscheiden. Daher haben sich die Autoren für einen Regression Calibration Ansatz entschieden, welcher im Folgenden genauer betrachtet wird.

7.1.1 Regression Calibration in der Studie

Regression Calibration Ansatz Zunächst wird angenommen, dass sich der health outcome auf folgende Weise beschreiben lässt

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \{\psi_1 Z_{it}^1 + \dots + \psi_k Z_{it}^k\} + \gamma_i + \epsilon_{it} \quad . \quad (29)$$

Y_{it} beschreibt die Ein-Sekunden-Kapazität, X_{it} die wahre Feinstaubbelastung, Z_{it}^k sind weitere erklärende Variablen und γ_i ist ein subjekt-spezifischer zufälliger Intercept. Der Index i bezeichnet das jeweilige Individuum und t die Zeit. Wie in dem Beispiel aus Kapitel 5.2 ist β_1 der Parameter von Interesse, da dieser den Effekt der wahren Feinstaubbelastung auf die Gesundheit beschreibt. Jedoch kann die wahre Feinstaubbelastung nicht direkt gemessen werden, weswegen diese modelliert werden muss

$$X_{it} = \alpha_0 + \alpha_1 W_t + \phi_i W_t + \omega_{it} \quad . \quad (30)$$

Das W_t entspricht wie in den vorherigen Kapiteln dem Surrogat für X_{it} , also in diesem Fall dem an Außenstellen gemessenen Feinstaub. Der Effekt, der über alle Individuen gleich ist, wird mit α bezeichnet, wohingegen ϕ_i einen zufälligen subjekt-spezifischen slope Parameter darstellt. Da auch die persönliche Feinstaubbelastung geschätzt wird, sollte ein einfaches Fehlermodell berücksichtigt werden. Dieses ergibt sich zu $X_{it}^* = X_{it} + u_{it}$, wobei sich also in Gleichung (30) lediglich der Fehlerterm zu $\omega'_{it} = \omega_{it} + u_{it}$ verändert.

Das Modell für die persönliche Feinstaubbelastung (30) kann in das health model (29) eingesetzt werden. Dann ergibt sich das erweiterte health model

$$Y_{it} = [\beta_0 + \beta_1 \alpha_0] + [\beta_1 \alpha_1] W_t + [\beta_1 \phi_i] W_t + \{\psi_1 Z_{it}^1 + \dots + \psi_k Z_{it}^k\} + \gamma_i + [\beta_1 \omega_{it} + \epsilon_{it}] \quad . \quad (31)$$

Zunächst einmal fällt auf, dass in Gleichung (31) nun ein subjekt-spezifischer slope Parameter über Gleichung (30) eingearbeitet wurde, obwohl das ursprüngliche health model (29) keinen solchen Parameter berücksichtigt hatte. Auch hier kann wieder $\beta_1 \alpha_1 = \delta_1$ definiert werden. Denn letztendlich ist der Gesundheitseffekt der wahren Feinstaubbelastung β_1 von Interesse, welcher nun analog zu dem Beispiel

aus Kapitel 5.2 mit

$$\hat{\beta}_1 = \hat{\delta}_1 / \hat{\alpha}_1$$

leicht berechnet werden kann, da sowohl $\hat{\alpha}_1$ aus Modell (30) als auch $\hat{\delta}_1$ aus Modell (31) bestimmt werden können.

Vorteile von Regression Calibration Es wird angenommen, dass der Feinstaub im Haus ein fester Anteil des Feinstaubes von außerhalb ist. Da die Menschen aber den Großteil ihrer Zeit im Haus verbringen, entsteht ein konzentrationsabhängiger Messfehler, weil die Differenz zwischen dem an Außenstellen gemessenen Wert und wahren Wert im Haus bei stärkerer Feinstaubkonzentration größer wird. Diese Art von Messfehler kann mit dem Regression Calibration Ansatz vermieden werden, da der naive Schätzer durch α_1 geteilt wird.

Wird eine mit Messfehler gemessene Variable als Prädiktor in einem Modell verwendet, entsteht ein Bias Richtung Null, der sogenannte *attenuation bias*. Dieses Problem kann mit Regression Calibration gelöst werden. Denn wird diese fehlerbehaftete Variable als abhängige Variable in einem Modell geschätzt, wird der Messfehler im Residuen-Fehlerterm absorbiert. Werden dann die geschätzten Werte der Variable im eigentlichen Modell verwendet, wird der *attenuation bias* vermieden. Dies gilt in vorliegender Studie vor allem für die Variable der persönlichen Feinstaubbelastung.

Des Weiteren liegen nur wenige Messungen für die persönliche Feinstaubbelastung vor, jedoch verhältnismäßig viele für den Bereich außerhalb des Hauses. Daher kann man meistens keine Modelle benutzen, die den persönlichen Feinstaub direkt als Prädiktor verwenden, weil die Aussagekraft zu gering ist. Durch den Regression Calibration Ansatz kann man aber eine sehr gut vorhersagbare Beziehung zwischen dem persönlichen Feinstaub und dem an Außenstellen gemessenen herstellen. Daher ist es nicht nötig mehr Messungen der persönlichen Belastung durchzuführen, sondern es können große Mengen an Messungen von Außenstellen sinnvoll für die Modellierung eingesetzt werden. Dies reduziert die für so eine Studie nötigen Kosten erheblich.

7.1.2 Verschiedene Methoden für die Schätzung

Wie oben erwähnt werden bei Strand u. a. (2007) generell zwei verschiedene Methoden betrachtet, um den health outcome zu schätzen. Die erste ist die positive matrix factorization. Diese ist ein Faktor–Analyse–Modell und versucht die verschiedenen Quellen des Feinstaubes zu identifizieren und zu trennen. Somit soll herausgefiltert werden, welchen Feinstaubquellen die Individuen zu welchen Anteilen ausgesetzt sind.

Die zweite Methode ist die Extrapolation. Bei dieser Methode werden einzelne Bestandteile des Feinstaubes untersucht, die ausschließlich oder hauptsächlich von Außenquellen stammen. Diese werden dann benutzt, um Werte für die persönliche Feinstaubbelastung zu extrapolieren. Die Methode ist wesentlich einfacher zu verstehen und anzuwenden als die PMF, jedoch gibt es mehr Unsicherheit in Bezug auf die Bestandteile, die betrachtet werden sollten. Daher werden verschiedene Parameter in das Modell mit einbezogen. Insgesamt werden bei Strand u. a. (2007) vier verschiedene Methoden der Extrapolation betrachtet.

Die erste Extrapolationsmethode betrachtet lediglich die Sulfatpartikel im Feinstaub. Es gilt approximativ

$$\frac{S_t}{W_t} \approx \lambda \frac{S_{it}}{X_{it}} \quad . \quad (32)$$

Dabei beschreibt S_t die Sulfatkonzentration in dem an Außenstellen gemessenen Feinstaub. S_{it} bezeichnet die persönliche Sulfatbelastung, λ ist ein Korrekturfaktor. Er beschreibt die durchschnittliche relative Differenz zwischen dem Anteil der Sulfate im Feinstaub außen und dem Anteil der Sulfate in der persönlichen Feinstaubumgebung. λ kann aus den beobachteten Daten geschätzt werden. Da aber wieder X_{it} von Interesse ist und berechnet werden soll, kann Gleichung (32) umgeschrieben werden zu

$$X_{it}^* = W_t \left(\frac{\hat{\lambda} S_{it}}{S_t} \right) \quad . \quad (33)$$

Ganz analog zu Formel (33) können die anderen Extrapolationsmethoden definiert werden. Die zweite Methode extrapoliert sowohl die Sulfat– als auch die

Eisenpartikel (I_t)

$$X_{it}^* = W_t \left(\frac{\hat{\lambda}_S S_{it} + \hat{\lambda}_I I_{it}}{S_t + I_t} \right) .$$

Die dritte Methode ist wie die zweite, nur dass noch ein Dummy für die Temperatur eingeführt wird. Es wird unterschieden zwischen Temperaturen unter 38 Grad Fahrenheit und Temperaturen darüber

$$X_{it}^* = W_t \left(\frac{\hat{\lambda}_{S,\text{temp}<38} S_{it} + \hat{\lambda}_{I,\text{temp}<38} I_{it}}{S_t + I_t} \right) ,$$

$$X_{it}^* = W_t \left(\frac{\hat{\lambda}_{S,\text{temp}\geq 38} S_{it} + \hat{\lambda}_{I,\text{temp}\geq 38} I_{it}}{S_t + I_t} \right) .$$

Die vierte und letzte Methode extrapoliert wiederum nur die Sulfate, bezieht jedoch keinen Korrekturfaktor mit ein. Es gilt also $\lambda = 1$.

7.1.3 Ergebnisse

Zunächst einmal werden die λ -Parameter betrachtet. Diese nehmen für die Sulfate stets einen Wert kleiner als eins an. Dies ist nicht verwunderlich, da die Sulfatpartikel eher klein und leicht sind und somit besser in die Häuser und Innenräume eindringen können. Daher ist der relative Anteil im Haus größer als der relative Anteil außerhalb des Hauses und λ wird Werte kleiner eins annehmen. Anders sieht es bei den Eisenpartikeln aus. Diese sind eher größer und schwerer, wodurch der relative Anteil im Haus kleiner ist als der relative Anteil außerhalb, weil die Partikel nicht leicht nach innen dringen können. Die Werte von λ sind demnach durchgehend größer als eins.

Interessant ist, dass für alle Partikel der Wert von λ näher an eins liegt, sobald wärmere Temperaturen vorherrschen. Dies ist vermutlich auf offene Fenster und Klimaanlage zurückzuführen, da diese die Außen- und Innenluft besser vermischen und somit die Anteile relativ gesehen wieder ausgeglichen werden.

Nachdem die Werte für λ berechnet wurden, kann die persönliche Feinstaubbelastung geschätzt werden. Als erstes soll die Verteilung des Anteils des gemessenen Feinstaubes betrachtet werden, dem die Individuen ausgesetzt sind (F_{PEX}). Das Augenmerk soll vor allem auf die letzte Spalte in Tabelle 6 gelegt werden. Dort

Estimation method	0	5	25	50	75	95	100	% > 1
PMF	0.07	0.27	0.41	0.54	0.78	1.28	11.05	12.1
Extrapolation 1	0	0.18	0.35	0.47	0.62	1.03	2.91	5.6
Extrapolation 2	0.05	0.23	0.38	0.50	0.63	0.94	2.18	3.9
Extrapolation 3	0.07	0.25	0.40	0.54	0.70	1.03	2.50	6.9
Extrapolation 4	0	0.23	0.45	0.61	0.80	1.33	3.75	13.4

Tabelle 6: Perzentile für F_{PEX}

sind die Prozentsätze der F_{PEX} -Werte abgetragen, welche die 100 % überschreiten. Dies ist theoretisch nicht möglich, da die Individuen höchstens dem gesamten Feinstaub ausgesetzt sein können, niemals aber mehr. Daher dient dies als erster Indikator dafür, welche Methode besser geeignet ist. Man sieht, dass die Methoden 1, 2 und 3 der Extrapolation am besten abschneiden.

Tabelle 7 zeigt zum Einen die Schätzwerte für die α_1 mit zugehörigen Standardfehlern in Klammern und zum Anderen die Residuenvarianzen. Man kann erkennen, dass die Schätzwerte für die α_1 eng zusammen liegen, wenn man Methode 4 der Extrapolation ausklammert. Weiter sieht man, dass auch die entsprechenden Standardfehler sehr ähnlich sind, wieder mit der Ausnahme von Methode 4. Auch weisen die Residuenvarianzen keine großen Unterschiede auf, wobei Methode 2 als die beste und Methode 4 bei weitem als die ungenaueste erscheinen.

Estimation method	$\hat{\alpha}_1$ (SE)	$\hat{\sigma}_{\omega'}^2$
PMF	0.51 (0.020)	9.7
Extrapolation 1	0.46 (0.019)	10.2
Extrapolation 2	0.47 (0.017)	6.4
Extrapolation 3	0.50 (0.019)	8.1
Extrapolation 4	0.59 (0.025)	17.0

Tabelle 7: Schätzungen für α_1 und Residuenvarianz

In Abbildung 5 sind nun die Schätzungen für β_1 abgetragen. In diesem Fall bedeutet *unadjusted*, dass keine der fünf Methoden benutzt wurde, sondern der Schätzwert der einfachen Regression Calibration verwendet wird. Die -1 % bedeuten, dass die Ein-Sekunden-Kapazität pro 10 μm mehr an Feinstaub um 1 % abnimmt, sich also verschlechtert. Man kann sehen, dass alle Methoden einen betragsmäßig größeren Schätzer liefern als die nicht-angepasste Schätzung. Jedoch werden auch die Konfidenzintervalle größer. Dies bedeutet, dass die Schätzungen

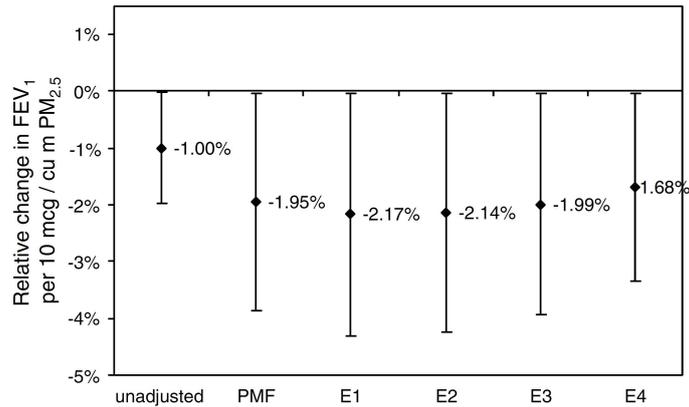


Abbildung 5: Schätzer für β_1

eine etwas größere Streuung aufweisen. Weiter sieht man, dass auch bei den Schätzungen für β_1 die Werte nicht weit auseinanderliegen. Als Hinweis sei gesagt, dass sich alle Werte für die Methode 4 entsprechend der Definition als lineare Transformationen der Methode 1 ergeben.

Die Autoren gehen letztendlich davon aus, dass die PMF und die Methoden 1–3 der Extrapolation generell geeignet sind, um die Analysen durchzuführen. Jedoch sollte darauf geachtet werden, dass nicht zu viele Parameter benutzt werden. Daher scheint es nicht immer sinnvoll einen Temperatur-Dummy aufzunehmen, da Methode 3 nicht genauer oder besser ist als Methode 2. Allerdings kann es nützlich sein, wenn man Messungen zu verschiedenen Jahreszeiten vergleicht. Dem Problem der F_{PEX} -Werte über Eins sollte man bereits vor der Analyse begegnen und nicht erst hinterher. Dies kann durch eine Nebenbedingung erfolgen, wenn von vornherein keine Werte über Eins zugelassen werden.

Die PMF ist im Allgemeinen schwieriger durchzuführen als die Extrapolation und es können zudem Probleme bei der analytischen Berechnung auftreten. Bei der Extrapolation treten jedoch gewisse Ungenauigkeiten auf, weswegen der Korrekturterm λ benutzt werden sollte. Doch auch bei der Schätzung von λ können Probleme auftreten. Vor allem wenn Fehler in der Spezifikation vorliegen, nimmt die Präzision ab und es können Fehlerterme mit Erwartungswert ungleich Null auftreten, was letztendlich die Modellannahmen verletzt. Daher bedarf es noch

weiterer Forschung, um noch bessere Methoden zur Schätzung des health outcomes zu entwickeln. An dieser Stelle sei nochmals daraufhingewiesen, dass Regression Calibration die Evaluierung der verschiedenen Methoden erst ermöglicht. Denn bevor unterschiedliche Methoden verglichen werden können, muss zunächst eine Beziehung zwischen der wahren Feinstaubbelastung X_{it} , dem Surrogat W_t und dem health outcome Y_t hergestellt werden.

7.2 Studie von Spiegelman

Auch in der Studie von Spiegelman (2010) wird Regression Calibration zur Behandlung von Messfehlern benutzt. Wie Blair u. a. (2007, Seite 205) schreiben, so geht auch Spiegelman (2010) davon aus, dass in epidemiologischen Studien die zwei Hauptprobleme das Confounding und die Misklassifikation der Exposition darstellen, wobei letzteres mit Abstand die größten Schwierigkeiten bereitet. Bias entstehe hauptsächlich auf Grund von Messfehlern und der damit verbundenen fehlerhaft spezifizierten Beziehung zwischen der wahren und der gemessenen Exposition. Zudem sei anzunehmen, dass in fast jeder epidemiologischen Studie die Exposition oder die Dosis nur mit Messfehler gemessen werden kann. Daher zeigt Spiegelman (2010) wie der Messfehlerproblematik in drei verschiedenen Situationen mit Regression Calibration begegnet werden kann. Diese drei Szenarien werden im folgenden genauer beschrieben.

7.2.1 Hauptstudie und externe Validierungsstudie

Das Design dieser Studie ist sehr ähnlich zu den in Kapitel 6 vorgestellten Simulationen. Hier allerdings wird die Methode der Regression Calibration an realen Daten angewendet. Als Datengrundlage dient *The Dutch Study of Respiratory Effects of Outdoor Air Pollution Constituents*. Die Hauptstudie wurde in den Jahren 1997 und 1998 in den Niederlanden durchgeführt und beinhaltet 2083 Schüler, deren Schule innerhalb von 400 Metern eines Freeways steht. Dabei wurde die Rußbelastung außerhalb der Schule gemessen (Janssen u. a., 2003). Die externe Validierungsstudie fand sechs Jahre später statt, wobei für 54 niederländische Kinder in vier Schulen Messungen der persönlichen NO_x -, $\text{PM}_{2.5}$ - und Rußbelastung vollzogen wurden (Van Roosbroeck u. a., 2007). Van Roosbroeck u. a. (2008) fassten dann beide Studien in einem Haupt-/Validierungsstudiendesign zusammen.

Als abhängige Gesundheitsvariablen dienen in dieser Untersuchung die binären Indikatoren für Keuchen, Bindehautentzündung, Phlegma und erhöhte Werte von Immunglobulin E, die jeweils den Wert 1 annehmen, falls ein Kind entsprechende Symptome bzw. Krankheiten aufweist. Diese Variablen wurden zusammen mit den Surrogaten der Exposition und weiteren erklärenden Variablen in der Hauptstudie gemessen. In der Validierungsstudie wurden dann die wahre persönliche Exposition, die Surrogate der Exposition und die weiteren erklärenden Variablen betrachtet. Für die Analyse der Daten wurde ein log-binomial Modell geschätzt. Dieses hat den Vorteil direkt das Prävalenz-Verhältnis (*prevalence ratio*) anzugeben und nicht das Prävalenz-Chancenverhältnis (*prevalence odds ratio*). Letzteres approximiert ersteres und liefert eher schwache Ergebnisse für den Fall, dass die Krankheit nicht selten vor kommt. Das Modell lautet wie folgt

$$P(Y = 1|X, Z_1, \dots, Z_s) = \exp(\beta_0 + \beta_1 X + \beta_{21} Z_1 + \dots + \beta_{2s} Z_s) \quad . \quad (34)$$

Die Variablen sind wie in den vorherigen Kapiteln definiert. Y ist dabei eine der vier oben definierten Gesundheitsvariablen und X ist die wahre persönliche Rußbelastung gemessen in $\mu\text{g}/\text{m}^3$. Die zusätzlichen erklärenden Variablen ($s = 1, \dots, 6$) sind in dieser Studie elterliches Rauchen, vorhandener Gasherd, vorhandener nicht ausgelassener Boiler, Geschlecht, Alter und ob Haustiere vorhanden sind. Gemäß der Standardverfahren kann das relative Risiko definiert werden als

$$\frac{P(Y = 1|X + \Delta, Z_1, \dots, Z_s)}{P(Y = 1|X, Z_1, \dots, Z_s)} = \exp(\beta_1 \Delta) \quad ,$$

wobei Δ eine wissenschaftlich sinnvolle Erhöhung der Exposition darstellt.

Entsprechend dem Regression Calibration Algorithmus muss nun ein Modell für die wahre Rußbelastung geschätzt werden, da diese in der Hauptstudie nicht beobachtet werden konnte. In diesem Fall ergibt sich das lineare Modell

$$X = \alpha_0 + \alpha_1 W + \alpha_{21} Z_1 + \dots + \alpha_{26} Z_6 + \epsilon \quad , \quad (35)$$

wobei W das Surrogat für die Exposition aus der Validierungsstudie darstellt. Als letzter Schritt der Regression Calibration muss X in Gleichung (34) mit der Schätzung für X aus Gleichung (35) ersetzt werden.

Bei Berechnung des Modells (35) konnte festgestellt werden, dass sowohl die Messung der Außenluft an Schulen als Surrogat als auch das elterliche Rauchen signifikanten Einfluss auf die persönliche Rußbelastung haben. Dabei war eine Korrelation zwischen dem Surrogat W und der wahren persönlichen Belastung X von 0.53 zu beobachten. Betrachtet man dann die Ergebnisse für das interessierende Modell (34), so konnte Spiegelman (2010) zeigen, dass die nicht bereinigte Schätzung des Prävalenz-Verhältnisses von Bindehautentzündung 2.18 beträgt. Wird aber der Regression Calibration Ansatz verfolgt und die Messfehler bereinigt, ergibt sich ein Prävalenz-Verhältnis von 5.06. Das Konfidenzintervall der Schätzung wird zwar wie erwartet etwas größer, doch muss man davon ausgehen, dass der Effekt der Rußbelastung auf das Risiko einer Bindehautentzündung um mehr als die Hälfte unterschätzt wurde, falls die Messfehler unberücksichtigt bleiben.

7.2.2 Mehrere Surrogate

In einigen Fällen ist es möglich, dass nicht nur ein einziges Surrogat für die wahre Exposition verfügbar ist, sondern mehrere Variablen existieren, die in einem gewissen Verhältnis zur eigentlichen Expositionsvariable stehen. Als Grundlage für die folgende Anwendung der Regression Calibration dient *The Six Cities Study of Respiratory Effects of Indoor Exposure to NO₂* mit internen Validierungsdaten. Die binär kodierte abhängige Variable Y nimmt den Wert Eins an, wenn innerhalb des letzten Jahres Symptome für Erkrankungen der unteren Atemwege festgestellt werden konnten. Dabei wird die Auswirkung der NO₂-Belastung auf diese Variable betrachtet.

Zur Analyse wird ein logistisches Modell angenommen

$$P(Y_i = 1) = \beta_0 + \beta_1 X_i + \beta_{21} Z_{1i} + \dots + \beta_{2s} Z_{si} \quad . \quad (36)$$

Die Variablen sind wie in den vorherigen Kapiteln definiert. X beschreibt nun die persönliche Exposition zu NO₂. Wieder ist β_1 der Parameter von Interesse und beschreibt das logarithmierte Odds-Ratio für einen Anstieg der Exposition um eine Einheit. Im nächsten Schritt der Regression Calibration muss nun wieder ein Modell für die wahre Exposition entwickelt werden. Dies geschieht analog zu Gleichung (35) mit dem Unterschied, dass r Surrogate berücksichtigt werden

müssen.

$$X_i = \alpha_0 + \alpha_{11}W_{1i} + \dots + \alpha_{1r}W_{ri} + \alpha_{21}Z_{1i} + \dots + \alpha_{2s}Z_{si} + \epsilon_1 \quad . \quad (37)$$

Wird dann X aus Gleichung (36) wieder mit der Schätzung von X aus Gleichung (37) substituiert, erhält man das Modell

$$P(Y_i = 1) = \delta_0 + \delta_{11}W_{1i} + \dots + \delta_{1r}W_{ri} + \delta_{21}Z_{1i} + \dots + \delta_{2s}Z_{si} \quad .$$

Analog zu Kapitel 5.2 können die einzelnen Schätzer für β ermittelt werden als

$$\hat{\beta}_{1j} = \frac{\hat{\delta}_{1j}}{\hat{\alpha}_{1j}} \quad \text{mit} \quad j = 1, \dots, r \quad .$$

Jedoch werden diese Schätzer noch mit Hilfe von inversen Varianz–Gewichten kombiniert. Denn so entsteht ein einzelner Schätzer $\hat{\beta}_1$ mit minimaler Varianz und der ursprünglichen physikalischen Einheit der Expositionsvariable. Tabelle 8 zeigt die einzelnen Surrogate mit ihren entsprechenden Gewichten.

Variable	Gewicht
Gasofen ohne Kontrollleuchte	0.18
Gasofen mit Kontrollleuchte	0.51
Ofenheizung	0.002
Ventilator	0.03
Holzofen	0.04
Anzahl der Räume im Haus	0.03
Petroleumheizung	0.20

Tabelle 8: Surrogate und entsprechende Gewichte

Führt man die Analyse mit Regression Calibration durch, erhält man ein Odds–Ratio von 1.60. Die einzelnen Surrogate haben dabei stark unterschiedliche Gewichte, welche sich in Summe zu Eins addieren müssen. Dieses Resultat liegt nahe an dem Odds–Ratio von 1.41, dem Ergebnis, wenn die Validierungsstudie einzeln berechnet wird. Werden beide Ansätze zusammengefügt, ändert sich das Odds–Ratio geringfügig zu 1.45, jedoch nimmt die Präzision zu. Dies kann an einem schmäleren Konfidenzintervall gesehen werden. Das Ergebnis lässt sich also in der Art interpretieren, dass ein Anstieg des NO₂ um 15 ppb (parts per billion) ei-

ne Erhöhung der Chance auf ein oder mehrere Symptome einer Erkrankung der unteren Atemwege um ca. 50 % zur Folge hat.

Der Regression Calibration Ansatz mit mehreren Surrogaten hat Vorteile gegenüber der Methode mit nur einem Surrogat, da alle im Datensatz vorhandenen Informationen genutzt werden können. Der Bias wird reduziert und auch ist der Verlust an Effizienz auf Grund einer höheren Varianz weniger stark ausgeprägt.

7.2.3 Heteroskedastische Varianz des Messfehlermodells

Expositionsvariablen wie NO₂ können schiefe Verteilungen aufweisen, wie Abbildung 6 zeigt. Dies kann in dem Messfehlermodell (Gleichung (35) und (37)) zu Heteroskedastizität führen, was die Modellannahmen verletzen würde. Daher kann der Regression Calibration Ansatz erweitert werden, um diesem Problem zu begegnen.

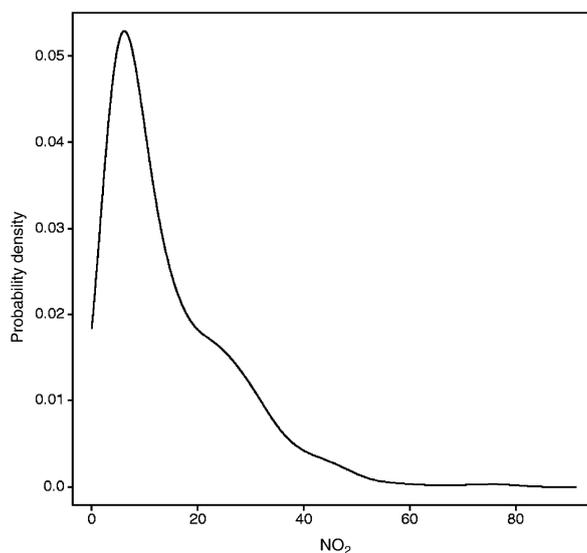


Abbildung 6: Verteilung der NO₂-Exposition

Zunächst einmal muss unter der Annahme, dass die Krankheit selten auftritt, definiert werden

$$P(Y = 1|X) = \frac{\exp(\beta_0 + \beta_1 X)Y}{1 + \exp(\beta_0 + \beta_1 X)} \approx \exp(\beta_0 + \beta_1 X)Y \quad .$$

Hinzu kommt, dass $f_2(X|W)$, eine allgemeine Formulierung des Messfehlermo-

dells, einer Normalverteilung folgt mit bedingtem Erwartungswert $\delta' + \alpha W$ und bedingter Varianz $\text{Var}(X|W) = h(W)\sigma^2$. Somit erhält man

$$\begin{aligned}
P(Y = 1|W) &= \int_X P(Y = 1|X)f_2(X|W) dX \\
&\approx \int_X \frac{\exp(\beta_0 + \beta_1 X - \frac{1}{2h(W)\sigma^2}(X - \delta' - \alpha W)^2)}{\sqrt{2\pi h(W)\sigma^2}} dX \\
&= \dots = \exp(\beta_0 + \beta_1 \delta' + \beta_1 \alpha W + \sigma^2 \beta_1^2 h(W)) \\
&= \exp(\beta_0^* + \beta_{11}^* W + \beta_{12}^* h(W)) \quad . \tag{38}
\end{aligned}$$

Die Berechnung des Schätzers $\hat{\beta}_{RCH}$ erfolgt daraufhin in drei Schritten. Als erstes muss in der Hauptstudie eine logistische Regression für Y auf (W, Z_1, \dots, Z_s) und $h(W)$ durchgeführt werden, um die Schätzer $\hat{\beta}_{11}^*$ und $\hat{\beta}_{12}^*$ mit ihren Varianzen zu erhalten. Als zweites muss in der Validierungsstudie eine gewichtete lineare Regression mit Gewichten $1/h(W)$ gerechnet werden. Hieraus erhält man die Schätzer $\hat{\sigma}^2$ und $\hat{\alpha}$. Im letzten Schritt erhält man $\hat{\beta}_{11}$ und $\hat{\beta}_{12}$, indem $\hat{\beta}_{11}^*$ und $\hat{\beta}_{12}^*$ für β mit Hilfe von $\hat{\sigma}^2$ und $\hat{\alpha}$ gelöst werden, wie in Gleichung (38) angedeutet. Somit kann man für den Schätzer $\hat{\beta}_{RCH}$ schreiben

$$\hat{\beta}_{RCH} = w_1 \hat{\beta}_{11} + (1 - w_1) \hat{\beta}_{12} \quad ,$$

wobei w_1 die asymptotisch minimalen Varianzgewichte darstellen.

Dieses Modell wurde von Spiegelman (2010) dann auf *The ACE Study of Occupational Exposure to Chemotherapeutic Agents* angewendet. Von Interesse war das Prävalenz–Chancenverhältnis zwischen akuten Gesundheitseffekten und Exposition zu Chemotherapeutika bei medizinischem Personal. Als wahre Expositionsvariable dient ein genau geführtes Medikamentenexpositions–Tagebuch und als Surrogat für dieses ist ein Fragebogen vorhanden. Die Korrelation zwischen beiden Variablen wurde auf ca. 0.70 geschätzt. Die Betrachtung der Residuen deutete eine moderate Heteroskedastizität an.

Um die oben beschriebene Erweiterung der Regression Calibration benutzen zu können, muss als erstes eine Form für den Ausdruck $h(W)$ gefunden werden.

Nach Durchsichtung der Klasse von Funktionen der Form $h(W) = (W + b)^p$ in der Validierungsstudie stellte sich heraus, dass sich als optimale Lösung $h(W) = W$ ergibt. Für diesen Ansatz stellte sich ein Odds-Ratio von 1.24 (95 % KI [1.05, 1.48]) ein. Wieder kann man erkennen, dass ohne den Regression Calibration Ansatz ein Bias Richtung Null entsteht, denn das unkorrigierte Odds-Ratio lag bei 1.08 (95 % KI [1.02, 1.15]). Jedoch ist auch hier die erhöhte Varianz an einem breiteren Konfidenzintervall zu sehen.

Interessant ist, dass das Ergebnis mit Berücksichtigung der Heteroskedastizität sehr nahe an dem Ergebnis der Regression Calibration unter Annahme von Homoskedastizität liegt. Spiegelman (2010) führte daraufhin eine Simulationsstudie durch und konnte feststellen, dass in allen Simulationen der normale RC-Schätzer einen geringeren Bias und eine geringere Varianz aufwies als der Schätzer unter Annahme von Heteroskedastizität. Sie führte dies darauf zurück, dass die zusätzliche Komplexität des Modells mehr Bias verursacht als die Bereinigung der Heteroskedastizität an Bias reduzieren kann. Daher scheint es in Situationen geringer bis mittlerer Heteroskedastizität durchaus angebracht das normale und einfachere Verfahren der Regression Calibration anzuwenden.

8 Ansätze in der Ökonometrie

In den vorangegangenen Kapiteln wurde beschrieben wie die Statistik in epidemiologischen Themenbereichen und die Statistik im Allgemeinen mit der Messfehlerproblematik umgehen. In diesem Zusammenhang stellt sich die Frage, ob in der Ökonometrie andere Modelle entwickelt wurden. Denn auch in der Ökonometrie sind Messfehler ein immer wieder auftretendes Problem. Nicht nur Daten der volkswirtschaftlichen Gesamtrechnung können teilweise nicht komplett erfasst werden, auch treten Messfehler bei Fragebögen oder wichtigen Kontrollvariablen auf. Im Folgenden wird daher erst sehr kurz das Konzept der *Generalized Method of Moments* (GMM) vorgestellt, bevor dann die Instrumentalvariablen-Schätzung erklärt wird. Diese gehört zur Familie der GMM und lässt sich mit Hilfe des Two Stage Least Squares-Ansatzes (TSLS-Ansatz) implementieren.

8.1 Generalized Method of Moments

Die Generalized Method of Moments (GMM) wurde von Hansen (1982) grundlegend als Erweiterung zur Momentenmethode definiert und dient seitdem nicht nur im Bereich der Messfehler als Standardkonzept in der Ökonometrie. Die GMM wird meist dort eingesetzt, wo auf Grund von unbekanntem Verteilungen Maximum-Likelihood-Schätzer nicht verfügbar sind oder nur mit sehr großem analytischen und rechentechnischen Aufwand geschätzt werden können.

Im normalen Modell der Momentenmethode definiert die Bedingung

$$g(\theta_0) = E[f(x_t, \theta_0)] = 0 \quad (39)$$

die Momente bzw. die orthogonale Bedingung des Modells. Dabei stellt θ ein $K \times 1$ Vektor von Parametern dar, $f(\cdot)$ einen R -dimensionalen Vektor von Funktionen und x_t mit $t = 1, \dots, T$ beinhaltet Modellvariablen. Existiert für diese Bedingung eine eindeutige Lösung genau dann wenn $\theta = \theta_0$, so ist das Modell identifiziert.

Das Gegenstück zum Erwartungswert aus Gleichung (39) für eine Stichprobe ist der Stichprobendurchschnitt

$$g_T(\theta) = \frac{1}{T} \sum_{t=1}^T f(x_t, \theta) \quad .$$

Gilt $R = K$ ist das Modell exakt identifiziert und man erhält die normalen Momentenmethode-Schätzer $\hat{\theta}_{MM}$ als Lösung von $g_T(\hat{\theta}_{MM}) = 0$. Ist das Modell jedoch überidentifiziert, gilt also $R > K$, gibt es im Allgemeinen keine Lösung für $g_T(\theta) = 0$. Hier schafft der GMM-Ansatz Abhilfe. Anstatt eine Lösung für $g_T(\theta) = 0$ zu suchen, soll $g_T(\theta)$ so klein wie möglich werden, d.h. der Abstand zur Null soll minimiert werden. Dieser Abstand wird über die folgende quadratische Form gemessen

$$Q_T(\theta) = g_T(\theta)' W_T g_T(\theta) \quad ,$$

wobei W_T eine $R \times R$ symmetrische und positiv definite Gewichtsmatrix darstellt. Nach Berechnung der optimalen Gewichtsmatrix (Hansen, 1982) ergibt sich der GMM-Schätzer in Abhängigkeit der Gewichtsmatrix zu

$$\hat{\theta}_{GMM}(W_T) = \arg \min_{\theta} \{g_T(\theta)' W_T g_T(\theta)\} \quad .$$

Dieser Schätzer ist in der Tat konsistent, wenn die Momenten-Bedingung korrekt ist. Denn es gilt

$$\hat{\theta}_{GMM}(W_T) \rightarrow \theta_0 \quad \text{wenn } T \rightarrow \infty \quad ,$$

für jedes positiv definite W_T . Intuitiv kann man sagen, dass $g_T(\theta)$ gegen $g(\theta)$ konvergiert, falls das Gesetz der großen Zahlen gilt. Da $\hat{\theta}_{GMM}(W_T)$ die Distanz von $g_T(\theta)$ zur Null minimiert, wird es ein konsistenter Schätzer der Gleichung $g(\theta_0) = 0$ sein.

8.2 IV-Schätzung

Der GMM-Ansatz kann in sehr vielen Bereichen eingesetzt werden und benötigt keine kompletten Verteilungsannahmen wie die ML-Methode. Als sehr bekanntes Beispiel der GMM ist das Verfahren der Instrumentalvariablen-Schätzung (IV-Schätzung) zu nennen. Diese wird im Folgenden genauer beschrieben. Dabei bezieht sich das Kapitel auf Stock u. Watson (2007, Kapitel 12) soweit nichts anderes angegeben wird.

8.2.1 Allgemeines Konzept

Mit dem Ansatz der IV-Schätzung lassen sich Endogenitätsprobleme von erklärenden Variablen behandeln. Ein Endogenitätsproblem liegt vor, wenn eine erklärende Variable mit dem Fehlerterm der Regression korreliert ist. Dadurch werden die Modellannahmen der Regression verletzt. Dies kann der Fall sein, wenn ein Confoundingproblem vorliegt, also eine Variable nicht in das Modell aufgenommen wurde, obwohl sie relevant ist. Dadurch entsteht der sogenannte Omitted Variable Bias. Genauso tritt das Problem der Endogenität auf, wenn ein kausaler Zusammenhang von der erklärenden Variable auf die abhängige Variable besteht und zugleich auch ein kausaler Zusammenhang der abhängigen Variable auf die erklärende Variable. Hier liegt ein Problem von simultaner Kausalität vor. Eine dritte Möglichkeit von Endogenitätsproblemen tritt bei Messfehlern auf, da die erklärende Variable über den Messfehler mit dem Fehlerterm korreliert ist. Zum ersten Mal findet der IV-Ansatz im Anhang bei Wright (1928) Erwähnung. Seitdem wurde der Ansatz weiter verfolgt und entwickelte sich zu einem Standardverfahren in der Ökonometrie. Der IV-Schätzer gehört zur Familie der GMM, welche zuvor näher

beschrieben wurde.

Die generelle Idee des IV-Schätzers ist es den Teil der erklärenden Variable zu isolieren, welcher nicht mit dem Fehlerterm korreliert ist um so konsistente Schätzer zu erhalten. Es soll also die Information über die Variation bzw. Bewegung in der erklärenden Variable herausgefiltert werden, die keine Korrelation mit dem Fehlerterm aufweist. Hierzu bedient man sich einer oder mehrerer Instrumentalvariablen, welche zwar mit der erklärenden Variable korrelieren, nicht aber mit dem Fehlerterm (Stock u. Watson, 2007, Seite 421). Daraus leiten sich auch die zwei Bedingungen für ein valides Instrument ab. Wird eine einfache lineare Regression betrachtet

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \quad , \quad (40)$$

wobei Y die abhängige Variable, X die mit Messfehler gemessenen endogene Variable und ϵ den Fehlerterm darstellen, dann muss für ein valides Instrument W gelten

- Instrument-Relevanz: $\text{Cov}(W_i, X_i) \neq 0$
- Instrument-Exogenität: $\text{Cov}(W_i, \epsilon_i) = 0$.

Erfüllt ein Instrument das Kriterium der Relevanz, dann besteht eine Beziehung zwischen dem Instrument und der Variation in X . Erfüllt es zudem die Bedingung der Exogenität, so ist auch die Variation von X exogen, die durch die Instrumentalvariable abgedeckt wird. Diese exogene Variation kann dann benutzt werden, um den Koeffizienten β_1 zu schätzen. Zusammengefasst darf ein Instrument W also Y nur indirekt über dessen Beziehung zu X beeinflussen.

8.2.2 Two Stage Least Squares

Natürlich kann das Instrument nicht einfach in die eigentliche Gleichung (40) substituiert werden. Daher wird die Instrumentalvariable über den sogenannten Two Stage Least Squares-Schätzer (TSLS) implementiert. Wie der Name schon sagt, besteht dieser Ansatz aus zwei Stufen, sehr ähnlich der Regression Calibration. In der ersten Stufe wird das endogene X in zwei Teile gespalten, einen unproblematischen und einen problematischen Teil. Hierzu wird eine Beziehung zwischen X und W hergestellt

$$X_i = \alpha_0 + \alpha_1 W_i + u_i \quad . \quad (41)$$

Der unproblematische Teil ist $\alpha_0 + \alpha_1 W_i$. Dieser ist exogen und nicht mit ϵ_i korreliert, da W exogen und unkorreliert mit ϵ_i ist. Der problematische Teil ist u_i und stellt den Teil von X_i dar, welcher mit ϵ_i korreliert ist. Im Weiteren wird also nur der unproblematische Teil berücksichtigt, um dem Endogenitätsproblem zu entgehen. Dies geschieht, indem man Gleichung (41) mit einer einfach linearen Regression schätzt. Man erhält

$$\hat{X}_i = \hat{\alpha}_0 + \hat{\alpha}_1 W_i \quad . \quad (42)$$

Der zweite Schritt der TSLS-Schätzung besteht darin das geschätzte \hat{X}_i aus Gleichung (42) in die Gleichung (40) einzusetzen. Somit führt man dann eine Regression nur mit dem exogenen Anteil von X_i aus und vernachlässigt den endogenen Anteil. Am Ende erhält man die konsistenten Schätzer $\hat{\beta}_0^{TSLS}$ und $\hat{\beta}_1^{TSLS}$. Für einen Beweis, dass $\hat{\beta}_i^{TSLS}$ tatsächlich konsistent ist siehe Stock u. Watson (2007, Seiten 428–429). Analog zur Notation in Kapitel 5.2 müssten diese Schätzer eigentlich mit $\hat{\delta}_i$ bezeichnet werden. Jedoch stellen sie die finalen Schätzer des Modells dar, welche den Effekt von X auf Y quantifizieren. Daher werden sie im Folgenden dennoch mit $\hat{\beta}_i^{TSLS}$ bezeichnet.

Es fällt auf, dass der TSLS-Ansatz dem der Regression Calibration sehr ähnlich ist. Vom Algorithmus her laufen sie annähernd identisch ab. Jedoch wurde in Kapitel 5.2 gezeigt, dass sich der RC-Schätzer als δ/α berechnen lässt. Dieser letzte Schritt wird bei der IV-Schätzung nicht berücksichtigt. Denn hier kommt der Unterschied zwischen einem Instrument und einem Surrogat zum Tragen. Das Surrogat ist ein Ersatz für eine mit Messfehler gemessene Variable, wohingegen das Instrument in der Tat nur ein Instrument ist, welches dazu benutzt wird die endogene Variable in die zwei Komponenten (problematisch / unproblematisch) aufzuteilen. Dabei schreibt die Bedingung für ein valides Instrument vor, dass es mit der abhängigen Variable Y nur über den Zusammenhang mit X in Verbindung steht, also nicht direkt mit Y korreliert ist. Das Surrogat hingegen ist meistens auch mit Y korreliert, da es genau wie die endogene Variable den interessierenden Sachverhalt zu messen versucht. Ein Instrument stellt diesen Anspruch nicht. Daher genügt bei der IV-Methode der Schätzer $\hat{\beta}_1^{TSLS}$, welcher sich direkt aus dem eigentlichen Modell (40) mit \hat{X}_i statt X_i ergibt.

Natürlich lässt sich die IV-Schätzung auch anwenden, falls mehrere Instrumente

und mehrere endogene Variablen vorliegen. Hierbei werden alle endogenen Variablen X im ersten Schritt auf alle Instrumente W und exogenen Variablen Z regressiert. Auch dies ist einer der großen Unterschiede zur Regression Calibration. Ein Beispiel für den Fall, wenn mehrere Surrogate vorhanden sind, wurde in Kapitel 7.2.2 gegeben. Im zweiten Schritt werden dann alle endogenen X_i durch ihre Schätzungen aus dem ersten Schritt ersetzt. Dabei ist jedoch zu beachten, dass die Koeffizienten exakt oder überidentifiziert sein müssen. Dies bedeutet, dass mindestens genauso viele oder mehr Instrumente wie endogene Variablen vorhanden sein müssen. Hinzu kommt, dass die Bedingungen für valide Instrumente anzupassen sind. Die Bedingung der Instrument-Relevanz muss insoweit ergänzt werden, dass perfekte Multikollinearität zwischen $(X_1, \dots, X_k, Z_1, \dots, Z_r)$ ausgeschlossen wird. Die Instrument-Exogenität wird lediglich dahingehend erweitert, dass jedes W_1, \dots, W_m nicht mit dem Fehlerterm ϵ korreliert sein darf.

8.2.3 Probleme der IV-Schätzung

Die Voraussetzung für eine erfolgreiche IV-Schätzung ist es also Instrumentalvariablen zu finden, welche die zwei Kriterien für valide Instrumente erfüllen. Dies ist jedoch in der Praxis nicht immer leicht. Daher ist die größte Gefahr, dass man schwache Instrumente findet und mit diesen die Berechnungen durchführt. Schwache Instrumente sind gemäß ihrer Definition nur gering mit der endogenen Variable korreliert und führen zu inkonsistenten und teilweise sinnlosen Schätzungen. Daher ist es von größter Bedeutung gute Instrumentalvariablen zu identifizieren. Dabei dienen immer die zwei Bedingungen für valide Instrumente als die Kriterien, die überprüft werden müssen. Doch auch dies ist in der praktischen Anwendung oftmals nicht ohne Weiteres möglich.

Als erstes wird die Bedingung der Instrument-Relevanz betrachtet. Ist diese nicht oder nur in geringem Umfang gegeben, wird von schwachen Instrumenten gesprochen. Will man zum Beispiel den Effekt der Begabung eines Menschen auf sein Einkommen berechnen, tritt das Problem auf, dass man die Begabung nicht messen kann. Als mögliches Instrument könnte man sich die Ergebnisse eines IQ-Tests vorstellen. Je höher der IQ-Wert, desto größer ist die Begabung. Jedoch umfasst ein IQ-Test nur einen geringen Teil der Begabung und lässt zum Beispiel körperliche Aspekte außen vor. Der IQ-Test ist also lediglich sehr gering mit der

Begabung korreliert und kann als schwaches Instrument identifiziert werden.

Durch schwache Instrumente entsteht nicht nur das Problem, dass die Schätzung inkonsistent und sinnlos werden kann, sondern auch dass Modellannahmen verletzt werden. Liegt ein schwaches Instrument vor, ist $Cov(W_i, X_i)$ sehr gering und nahe Null. Dies führt dazu, dass der Schätzer $\hat{\beta}_1^{TOLS}$ erstens inkonsistent wird und zweitens auch asymptotisch nicht normalverteilt ist. Dies ist aber eine Voraussetzung der Modellannahmen für die Anwendung des TOLS-Ansatzes (Stock u. Watson, 2007, Seite 440). In diesem Zusammenhang stellt sich natürlich die Frage, wie schwache Instrumente erkannt werden können. Für den Fall, dass lediglich eine endogene Variable vorliegt, aber mehrere Instrumente, gibt es die Faustregel, dass die F-Statistik der Schätzung in Stufe 1 größer als 10 sein sollte (Stock u. Watson, 2007, Seiten 440–441). Für den Fall mehrerer endogener Variablen gibt es eine derartige Faustregel nicht. Hier muss man sich größtenteils auf sein persönliches Urteilsvermögen verlassen.

Als zweites wird die Bedingung der Instrument-Exogenität betrachtet. Es wird vorausgesetzt, dass ein Instrument nicht mit dem Fehlerterm der eigentlichen Regression korreliert ist. Dieses Kriterium kann jedoch nur in bestimmten Fällen und dann nur ungenau statistisch getestet werden. Sind die Koeffizienten exakt identifiziert, ist ein Test auf Exogenität nicht möglich. In einer solchen Situation muss man auf Expertenwissen oder eigene Überlegungen zurückgreifen. Es sind plausible Argumente nötig, um festzustellen, ob in dem vorliegenden Problem die Instrumentalvariablen in der Tat exogen sind oder ob berechtigte Zweifel an dieser Bedingung bestehen. Letztendlich läuft es immer auf eine persönliche Entscheidung hinaus.

Anders sieht es in dem Fall aus, wenn die Koeffizienten überidentifiziert sind. Hier lässt sich der sogenannte Test der überidentifizierenden Restriktionen (overidentifying restrictions test) durchführen. Liegen mehr Instrumente als endogene Variablen vor, wird im Rahmen dieses Tests für jede Instrumentalvariable einzeln der jeweilige IV-Schätzer ermittelt. Diese Schätzer werden dann miteinander verglichen. Weichen sie zu stark von einander ab, liegt vermutlich ein Problem in den Instrumenten vor und es ist davon auszugehen, dass eines oder mehrere Instrumente die Exogenitätsbedingung nicht erfüllen. Der Test beruht auf der Annahme, dass die einzelnen Instrumente zwar verschiedene aber sehr ähnliche Ergebnisse erzielen, da sie letztendlich mit dem gleichen Vorgehen das gleiche Ziel erreichen

wollen.

Zusammenfassend bleibt zu sagen, dass die IV-Schätzung als Standardmethode in der Ökonometrie vielfache Verwendung findet. Die Vorgehensweise über den TSLS-Ansatz ist intuitiv logisch und hat große Ähnlichkeiten mit dem Ansatz der Regression Calibration. Dennoch treten bei der Suche nach validen Instrumenten große Probleme auf. Nicht nur, dass zuerst mögliche Instrumente gefunden werden müssen, auch die Überprüfung der Bedingungen für valide Instrumente ist mit einer gewissen Ungenauigkeit verbunden. Die größte Einschränkung dabei besteht darin, dass man sich am Ende häufig auf seine eigene Meinung und sein persönliches Urteilsvermögen verlassen muss. Hat man jedoch valide Instrumente gefunden, liefert die IV-Schätzung schnelle und leicht zu interpretierende Ergebnisse. So schafften es Analysen mit Hilfe der IV-Schätzung sogar in das *Quarterly Journal of Economics*, die älteste und eine der angesehensten wirtschaftswissenschaftlichen Fachzeitschriften (Levitt, 1996; Angrist u. Krueger, 1991).

9 Abschließende Bemerkungen

Vorliegende Arbeit hat sich mit der Messfehlerproblematik bei epidemiologischen Studien in Bezug auf Luftschadstoffe beschäftigt. Dabei ist zuerst einmal wichtig zu wissen, welche Luftschadstoffe untersucht und analysiert werden sollen, da letztendlich die Wirkung der Schadstoffe auf den menschlichen Körper oder die Umwelt von Interesse ist. Denn schon bei der physikalischen Messung unterscheiden sich die einzelnen Schadstoffe sehr in Hinblick auf die Messgeräte und die Methodik. Bereits hier entstehen die ersten Messfehler in einer Studie, da keine Messgeräte oder Messstationen existieren, die stets Ergebnisse ohne Fehler liefern können. Diese Art von Messfehler wird zu meist als klassischer Messfehler interpretiert. Hinzu kommen dann Messfehler auf Grund des Studiendesigns oder anderer Faktoren.

Es wurde gezeigt, dass es für die Behandlung und Auswirkung der Fehler von großer Bedeutung ist, diese vom ihrem Typ her richtig als klassischen oder Berkson-Messfehler einzuordnen. Im Fall eines klassischen additiven Messfehlers wird sich stets ein Bias Richtung Null einstellen, d.h. der wahre Effekt wird unterschätzt. Anders verhält es sich bei einem additiven Berkson-Messfehler, da die-

ser keinen Bias produziert. Dennoch sollte auch letzterer nicht unberücksichtigt bleiben, da sich die Varianzen und Konfidenzintervalle verändern und somit zum Teil falsche Schlüsse beim Testen gezogen werden könnten. Zudem ist es in den meisten Situationen von Vorteil, wenn multiplikative Messfehler durch eine log-Transformation in additive verwandelt werden, da sich diese in Modellen wesentlich leichter handhaben lassen.

Als mögliches Modell um die Messfehler behandeln zu können, wurde die Methode der Regression Calibration genauer vorgestellt. Diese bietet einen recht simplen und intuitiv logischen Ansatz, indem statt den nicht vorhandenen wahren Werten die zuvor geschätzten Werte der Expositionsvariable im eigentlichen Modell benutzt werden. In den meisten Fällen lässt sich dieses Verfahren ohne großen Aufwand implementieren und liefert schnell erste Ergebnisse. Des Weiteren konnte festgestellt werden, dass dieses Verfahren trotz der Einfachheit in vielen Situationen gute Resultate liefert. Zwar scheinen die Maximum-Likelihood-Schätzer für den Fall von internen Validierungsdaten genauso wie die Schätzer der Multiple Imputation etwas weniger Bias aufzuweisen wie die Regression-Calibration-Schätzer, doch können letztere im Fall von externen Validierungsstudien zum Teil bessere Ergebnisse liefern. Vor allem in nicht extremen Szenarien bietet sich der Regression Calibration Ansatz an, da dieser schnell und mit weniger Aufwand als die anderen beiden Methoden modelliert werden kann. Zudem muss die Methode der Multiple Imputation für externe Validierungsstudien neu angepasst werden, was weitere Komplexität einführen würde.

Nicht nur die Methode der Maximum Likelihood kann flexibel angepasst werden. Auch Regression Calibration ist nicht zwangsläufig auf Validierungsdaten angewiesen ist. In Kapitel 5.3.2 wurde eine Möglichkeit vorgestellt, um Schätzer aus Wiederholungsmessungen zu erhalten. Zudem funktioniert der Ansatz auch, wenn ein unverzerrtes Instrument der eigentlichen Expositionsvariable vorhanden ist. Dies und die Tatsache, dass der Ansatz fast beliebig erweitert und angepasst werden kann, macht Regression Calibration zu einer sehr flexibel verwendbaren Methode, um Messfehler zu behandeln. Die Einfachheit und das breite Einsatzspektrum macht Regression Calibration auch in der Praxis sehr beliebt. Anhand von zwei Studien wurde gezeigt, wie der Ansatz erfolgreich angewendet wird und der Bias auf Grund von Messfehlern reduziert werden kann.

Zuletzt wurden noch Ansätze zur Messfehlerbehandlung aus der Ökonometrie

vorgestellt. Als Standardkonzept wurde die Generalized Method of Moments betrachtet. Diese hat gegenüber der ML-Methode den Vorteil, dass keine exakte Kenntnis der Verteilungen notwendig ist und sie daher in ungewissen Situationen oft schnellere und bessere Ergebnisse produziert. Jedoch bleibt der Maximum-Likelihood-Ansatz am besten, falls die Verteilungen bekannt sind. Als mögliches Anwendungsbeispiel der GMM wurde dann die Instrumentalvariablen-Schätzung behandelt. Diese hat sehr große Ähnlichkeiten mit dem Ansatz der Regression Calibration, unterscheidet sich aber letztendlich bei der Schätzung der Parameter. Generell ist auch das Prinzip eines Instruments von dem eines Surrogats verschieden, wie es bei Regression Calibration verwendet wird. Insgesamt konnte aber gezeigt werden, dass auch in der Ökonometrie bzw. der wirtschaftswissenschaftlichen Forschung die Messfehlerproblematik bekannt ist und behandelt werden muss, genauso wie in epidemiologischen Themengebieten.

Zusammenfassend bleibt festzuhalten, dass bei epidemiologischen Fragestellungen stets mit Messfehlern zu rechnen ist. Dabei treten sowohl klassische als auch Berkson-Messfehler auf, wobei ersterer auf Grund des Bias wesentlich stärkere Auswirkungen auf die Schätzung hat. Methoden wie Maximum Likelihood oder Regression Calibration bieten Möglichkeiten sich der Messfehlerproblematik anzunehmen, wobei sie bei weitem nicht die einzig möglichen Ansätze sind. Weiter ist davon auszugehen, dass ohne Berücksichtigung der Messfehler in jeder epidemiologischen Studie verzerrte Schätzer auftreten, wobei dieser Bias wahrscheinlich die größte Beschränkung der Validität solcher Studien darstellt. Daher schrieb Spiegelman (2010, Seite 159), dass genauso wie ein Modell routinemäßig für Confounding berichtigt wird, so sollte es auch routinemäßig für Messfehler berichtigt werden.

Anhang

I Tabellen zu Kapitel 6.5.2

Im Folgenden sind die beiden Tabellen 9 und 10 zu Kapitel 6.5.2 abgetragen. Tabelle 9 beschreibt die genauen Ergebnisse zu den Methoden der internen Validierungsstudie, Tabelle 10 die Ergebnisse der externen Validierungsstudie. Dabei steht *Mod.* für das entsprechende Modell, *Koef.* für die Koeffizienten, die betrachtet werden, *kompl. Daten* für die Ergebnisse des Goldstandards, wenn alle Daten verfügbar sind, *ML* für Maximum Likelihood, *RC* für Regression Calibration und *MI* für Multiple Imputation. Es wurden 1000 Datensätze mit je 500 Beobachtungen simuliert. Für die Ergebnisse des Bias sowie der Standardabweichung (SD) wurden in Klammern jeweils die entsprechenden Standardfehler angegeben.

In Tabelle 10 sind in der letzten Spalte die Ergebnisse der naiven Multiple Imputation angegeben. Diese bezeichnet die Ergebnisse der Multiple Imputation, wenn sie statt für interne Validierungsdaten analog für externe Validierungsdaten verwendet wird. Wie Messer u. Natarajan (2008, Seite 13) schreiben, funktioniert dieser Ansatz sehr schlecht und sollte so auch nicht verwendet werden. Stattdessen sollten für externe Validierungsdaten auch die fehlenden Y -Werte mit Multiple Imputation geschätzt werden (siehe Ende Kapitel 6.4).

Mod.	Koef.		kompl. Daten	ML	RC	MI
1	β_1	Bias	0.012(0.006)	0.037(0.008)	0.03(0.008)	0.001(0.008)
1	β_1	SD	0.18(0.003)	0.261(0.004)	0.265(0.004)	0.262(0.004)
1	β_2	Bias	0.012(0.004)	0.021(0.005)	0.022(0.006)	0.007(0.005)
1	β_2	SD	0.119(0.002)	0.169(0.003)	0.175(0.003)	0.167(0.003)
1	β_3	Bias	0.007(0.005)	0(0.006)	0.018(0.008)	0.016(0.006)
1	β_3	SD	0.166(0.003)	0.199(0.003)	0.242(0.004)	0.2(0.003)
2	β_1	Bias	0.016(0.006)	0.03(0.008)	0.027(0.008)	-0.002(0.008)
2	β_1	SD	0.179(0.003)	0.248(0.004)	0.252(0.004)	0.248(0.004)
2	β_2	Bias	0.008(0.005)	0.016(0.007)	0.016(0.007)	0.003(0.007)
2	β_2	SD	0.158(0.002)	0.231(0.004)	0.234(0.004)	0.233(0.004)
2	β_3	Bias	0.007(0.004)	0.01(0.004)	0.015(0.006)	0.008(0.004)
2	β_3	SD	0.118(0.002)	0.134(0.002)	0.178(0.003)	0.133(0.002)
3	β_1	Bias	0.009(0.006)	0.022(0.008)	0.023(0.008)	0(0.008)
3	β_1	SD	0.181(0.003)	0.239(0.004)	0.262(0.004)	0.245(0.004)
3	β_2	Bias	0.009(0.004)	0.016(0.005)	0.018(0.006)	0.009(0.005)
3	β_2	SD	0.118(0.002)	0.145(0.002)	0.176(0.003)	0.151(0.002)
3	β_3	Bias	0.014(0.005)	0.009(0.006)	0.022(0.007)	0.019(0.006)
3	β_3	SD	0.16(0.003)	0.183(0.003)	0.234(0.004)	0.188(0.003)
4	β_1	Bias	0.032(0.006)	0.038(0.007)	0.039(0.008)	0.015(0.008)
4	β_1	SD	0.184(0.003)	0.236(0.004)	0.262(0.004)	0.239(0.004)
4	β_2	Bias	-0.003(0.005)	0(0.007)	0.011(0.008)	-0.002(0.007)
4	β_2	SD	0.168(0.003)	0.216(0.003)	0.239(0.004)	0.225(0.004)
4	β_3	Bias	0.001(0.004)	0.001(0.004)	0.008(0.006)	0.002(0.004)
4	β_3	SD	0.122(0.002)	0.131(0.002)	0.177(0.003)	0.131(0.002)
5	β_1	Bias	0.055(0.01)	-0.096(0.012)	0.122(0.016)	-0.241(0.012)
5	β_1	SD	0.331(0.005)	0.388(0.006)	0.499(0.008)	0.393(0.006)
5	β_2	Bias	0.016(0.007)	0.038(0.01)	0.022(0.01)	-0.023(0.009)
5	β_2	SD	0.209(0.003)	0.309(0.005)	0.303(0.005)	0.276(0.004)
5	β_3	Bias	0.008(0.005)	0.001(0.006)	0.007(0.007)	-0.027(0.006)
5	β_3	SD	0.153(0.002)	0.186(0.003)	0.221(0.003)	0.18(0.003)

Tabelle 9: Ergebnisse für interne Validierungsstudie

Mod.	Koef.		kompl. Daten	ML	RC	naive MI
1	β_1	Bias	0.013(0.006)	0.24(0.04)	0.08(0.05)	-0.948(0.003)
1	β_1	SD	0.184(0.003)	1.37(0.02)	1.43(0.02)	0.085(0.001)
1	β_2	Bias	0.012(0.004)	0.09(0.03)	-0.06(0.02)	-0.321(0.002)
1	β_2	SD	0.119(0.002)	0.91(0.01)	0.69(0.01)	0.07(0.001)
1	β_3	Bias	0.006(0.005)	0.08(0.03)	-0.16(0.03)	0.552(0.006)
1	β_3	SD	0.162(0.003)	0.82(0.01)	1(0.02)	0.181(0.003)
2	β_1	Bias	0.006(0.006)	0.15(0.05)	-0.16(0.07)	-0.91(0.003)
2	β_1	SD	0.181(0.003)	1.51(0.02)	2.25(0.04)	0.087(0.001)
2	β_2	Bias	0.008(0.005)	0.15(0.05)	-0.07(0.07)	-0.296(0.003)
2	β_2	SD	0.164(0.003)	1.56(0.02)	2.16(0.03)	0.085(0.001)
2	β_3	Bias	0.013(0.004)	0.12(0.01)	-0.07(0.02)	0.072(0.005)
2	β_3	SD	0.123(0.002)	0.37(0.01)	0.52(0.01)	0.153(0.002)
3	β_1	Bias	0.015(0.006)	0.13(0.02)	-0.02(0.01)	-0.667(0.005)
3	β_1	SD	0.187(0.003)	0.6(0.01)	0.45(0.01)	0.145(0.002)
3	β_2	Bias	0.008(0.004)	0.01(0.01)	-0.03(0.01)	-0.146(0.004)
3	β_2	SD	0.121(0.002)	0.3(0.005)	0.26(0.004)	0.117(0.002)
3	β_3	Bias	0.007(0.005)	-0.02(0.01)	-0.04(0.01)	0.381(0.006)
3	β_3	SD	0.164(0.003)	0.37(0.01)	0.34(0.01)	0.193(0.003)
4	β_1	Bias	0.012(0.006)	0.12(0.02)	-0.06(0.02)	-0.571(0.004)
4	β_1	SD	0.185(0.003)	0.69(0.01)	0.5(0.01)	0.14(0.002)
4	β_2	Bias	0.004(0.005)	0(0.02)	-0.04(0.02)	-0.058(0.004)
4	β_2	SD	0.162(0.003)	0.58(0.01)	0.48(0.01)	0.135(0.002)
4	β_3	Bias	0.008(0.004)	0.01(0.01)	-0.04(0.01)	0.035(0.005)
4	β_3	SD	0.12(0.002)	0.22(0.004)	0.19(0.003)	0.157(0.002)
5	β_1	Bias	0.055(0.01)	-1.36(0.04)	-1.22(0.23)	-2.861(0.003)
5	β_1	SD	0.33(0.005)	1.32(0.02)	7.28(0.12)	0.098(0.002)
5	β_2	Bias	-0.002(0.006)	0.68(0.05)	-0.4(0.19)	-0.268(0.003)
5	β_2	SD	0.204(0.003)	1.65(0.03)	6.11(0.1)	0.094(0.001)
5	β_3	Bias	0.008(0.005)	0.18(0.01)	-0.25(0.05)	0.045(0.005)
5	β_3	SD	0.152(0.002)	0.43(0.01)	1.53(0.02)	0.147(0.002)

Tabelle 10: Ergebnisse für externe Validierungsstudie

Literatur

- [Angrist u. Krueger 1991] ANGRIST, J.D. ; KRUEGER, A.B.: Does Compulsory School Attendance Affect Schooling and Earnings? In: *The Quarterly Journal of Economics* 106 (1991), Nr. 4, S. 979–1014
- [Arndt 2001] ARNDT, S.: *Vergleich der Pferdehaltung in bäuerlich-ländlichen Kleinbetrieben mit derjenigen in hauptberuflichen, städtischen Pferdewirtschaftsbetrieben im Hinblick auf einen möglichen Zusammenhang mit Atemwegserkrankungen*, Justus-Liebig-Universität Gießen, Diss., 2001
- [Augustin u. a. 2008] AUGUSTIN, Thomas ; DÖRING, Alfred ; RUMMEL, David: Regression Calibration for Cox Regression under Heteroscedastic Measurement Error – Determining Risk Factors of Cardiovascular Diseases from Error-prone Nutritional Replication Data. In: SHALABH (Hrsg.) ; HEUMANN, C. (Hrsg.): *Recent Advances in Linear Models and Related Areas: Essays in Honour of Helge Toutenburg*. Physica-Verlag, Heidelberg, 2008. – ISBN 978-3-79082-063-8, S. 253–278
- [Blackwell 1947] BLACKWELL, D.: Conditional Expectation and Unbiased Sequential Estimation. In: *Annals of Mathematical Statistics* 18 (1947), Nr. 1, S. 105–110
- [Blair u. a. 2007] BLAIR, A. ; STEWART, P. ; LUBIN, J.H. ; FORASTIERE, F.: Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. In: *American Journal of Industrial Medicine* 50 (2007), S. 199–207
- [Carroll u. a. 2006] CARROLL, R.J. ; RUPPERT, D. ; STEFANSKI, L.A. ; CRAINCÉANU, C.M.: *Measurement Error in Nonlinear Models: A Modern Perspective*. 2nd Edition. Chapman and Hall/CRC, 2006. – ISBN 978-1-58488-633-4
- [Clayton u. a. 1993] CLAYTON, C.A. ; PERRITT, R.L. ; PELLIZZARI, E.D. ; THOMAS, K.W. ; WHITMORE, R.W. ; WALLACE, L.A. ; ÖZKAYNAK, H. ; SPENGLER, J.D.: Particle total exposure assessment methodology (PTEAM) study: distributions of aerosol and elemental concentrations in personal, indoor and outdoor air samples in a Southern Californian community. In: *Journal of Exposure Analysis and Environmental Epidemiology* 3 (1993), Nr. 2, S. 227–249
- [Dockery u. a. 1993] DOCKERY, D.W. ; POPE, C.A. ; XU, X. ; SPENGLER, J.D. ; WARE, J.H. ; FAY, M.E ; FERRIS, B.G. ; SPEIZER, F.E.: An association between air pollution and mortality in six U.S. Cities. In: *The New England Journal of Medicine* 329 (1993), Nr. 24, S. 1753–1759

- [Fick 1855] FICK, A.: Über Diffusion. In: *Analen der Physik* 170 (1855), Nr. 7, S. 59–86
- [Goldberg u. a. 2001] GOLDBERG, M.S. ; BURNETT, R. ; BAILAR, J.C. ; BROOK, J. ; BONVALOT, Y. ; TAMBLYN, R. ; SINGH, R. ; VALOIS, M.F.: The Association Between Daily Mortality and Short–Term Effects of Ambient Air Particle Pollution in Montreal, Quebec: 1. Nonaccidental Mortality. In: *Environmental Research* 86 (2001), Nr. 1, S. 12–25
- [Hansen 1982] HANSEN, L.P.: Large Sample Properties of Generalized Method of Moments Estimators. In: *Econometrica* 50 (1982), Nr. 4, S. 1029–1054
- [Heid u. a. 2004] HEID, I.M. ; KÜCHENHOFF, H. ; MILES, J. ; KREIENBROCK, L. ; WICHMANN, H.E.: Two dimensions of measurement error: Classical and Berkson error in residential radon exposure assessment. In: *Journal of Exposure Analysis and Environmental Epidemiology* 14 (2004), S. 365–377
- [Hoek u. a. 2001] HOEK, G. ; BRUNEKREEF, B. ; FISCHER, P. ; VAN WIJNEN, J.: The Association Between Air Pollution and Heart Failure, Arrhythmia, Embolism, Thrombosis, and Other Cardiovascular Causes of Death in a Time Series Study. In: *Epidemiology* 12 (2001), Nr. 3, S. 355–357
- [Janssen u. a. 2003] JANSSEN, N.A.H. ; BRUNEKREEF, B. ; VLIET, P. van ; AARTS, F. ; MELIEFSTE, K. ; HARSSEMA, H. ; FISCHER, P.: The Relationship between Air Pollution from Heavy Traffic and Allergic Sensitization, Bronchial Hyperresponsiveness, and Respiratory Symptoms in Dutch Schoolchildren. In: *Environmental Health Perspectives* 111 (2003), Nr. 12, S. 1512–1518
- [Janssen u. a. 1998] JANSSEN, N.A.H. ; HOEK, G. ; BRUNEKREEF, B. ; HARSSEMA, H. ; MENDINK, I. ; ZUIDHOF, A.: Personal sampling of particles in adults: relation among personal, indoor, and outdoor air concentrations. In: *American Journal of Epidemiology* 147 (1998), Nr. 6, S. 537–547
- [Janssen u. a. 2005] JANSSEN, N.A.H. ; LANKI, T. ; HOEK, G. ; VALLIUS, M. ; DE HARTOG, J.J. ; VAN GRIEKEN, R. ; PEKKANEN, J. ; BRUNEKREEF, B.: Associations between Ambient, Personal, and Indoor Exposure to Fine Particulate Matter Constituents in Dutch and Finnish Panels of Cardiovascular Patients. In: *Occupational and Environmental Medicine* 62 (2005), Nr. 12, S. 868–877
- [Keywood u. a. 1999] KEYWOOD, M.D. ; AYERS, G.P. ; GRAS, J.L. ; GILLETT, R.W. ; COHEN, D.D.: Relationships between size segregated mass concentration data and ultrafine particle number concentrations in urban areas. In: *Atmospheric Environment* 33 (1999), Nr. 18, S. 2907–2913

- [Lahl u. Steven 2005] LAHL, U. ; STEVEN, W.: Feinstaub – eine gesundheitspolitische Herausforderung. In: *Pneumologie* 59 (2005), Nr. 4, S. 704–714
- [Levitt 1996] LEVITT, S.D.: The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation. In: *The Quarterly Journal of Economics* 111 (1996), Nr. 2, S. 319–351
- [Levitt u. Dubner 2009] LEVITT, S.D. ; DUBNER, S.J.: *Superfreakonomics*. Penguin Books, London, 2009. – ISBN 978–0–141–03070–8
- [Lioy 1990] LIOY, P.J.: Assessing total human exposure to contaminants. In: *Environmental Science and Technology* 24 (1990), Nr. 7, S. 938–945
- [Messer u. Natarajan 2008] MESSER, K. ; NATARAJAN, L.: Maximum likelihood, multiple imputation and regression calibration for measurement error adjustment. In: *Statistics in Medicine* 27 (2008), Nr. 30, S. 6332–6350
- [Monn 2001] MONN, C.: Exposure assessment of air pollutants: a review on spatial heterogeneity and indoor/outdoor/personal exposure to suspended particulate matter, nitrogen dioxide and ozone. In: *Atmospheric Environment* 35 (2001), Nr. 1, S. 1–32
- [Morris 2007] MORRIS, E.: From Horse Power to Horespower. In: *Access: The Magazine of UCTC* 30 (2007), S. 2–9
- [Norris u. a. 1999] NORRIS, G. ; YOUNG-PONG, S.N. ; KOENIG, J.Q. ; LARSON, T.V. ; SHEPPARD, L. ; STOUT, J.W.: An Association Between Fine Particles and Asthma Emergency Department Visits for Children in Seattle. In: *Environmental Health Perspectives* 107 (1999), S. 489–493
- [Oezkaynak u. a. 1996] OEZKAYNAK, H. ; XUE, J. ; SPENGLER, J. ; WALLACE, L. ; PELLIZZARI, E. ; JENKINS, P.: Personal exposure to airborne particles and metals: results from the Particle Team Study in Riverside, C.A. In: *Journal of Exposure Analysis and Environmental Epidemiology* 6 (1996), S. 57–78
- [Ostro u. a. 2000] OSTRO, B.D. ; BROADWIN, R. ; LIPSETT, M.J.: Coarse and fine particles and daily mortality in the Coachella Valley, California: a follow-up study. In: *Journal of Exposure Analysis and Environmental Epidemiology* 10 (2000), S. 412–419
- [Ott 1982] OTT, W.R.: Concepts of human exposure to air pollution. In: *Environment International* 7 (1982), S. 179–196
- [Schafer 1997] SCHAFFER, J.L.: *Analysis of Incomplete Multivariate Data*. Chapman and Hall/CRC, Monographs on Statistics and Applied Probability, 72, 1997. – ISBN 978–0–41204–061–0

- [Seifert 1995] SEIFERT, B.: Validity criteria for exposure assessment methods. In: *Science of the Total Environment* 168 (1995), Nr. 2, S. 101–107
- [Shifa 2009] SHIFA, N.: *Estimation of QVF Measurement Error Models Using Empirical Likelihood Method*, Graduate College of Bowling Green State University, Diss., 2009
- [Spiegelman 2010] SPIEGELMAN, D.: Approaches to uncertainty in exposure assessment in environmental epidemiology. In: *Annual Review of Public Health* 31 (2010), S. 149–163
- [Stock u. Watson 2007] STOCK, J.H. ; WATSON, M.W.: *Introduction to Econometrics*. Second Edition. Pearson Education, 2007. – ISBN 0–321–44253–9
- [Strand u.a. 2007] STRAND, M. ; HOPKE, P. K. ; ZHAO, W. ; VEDAL, S. ; GELFAND, E. ; RABINOVITCH, N.: A study of health effect estimates using competing methods to model personal exposures to ambient PM_{2.5}. In: *Journal of Exposure Science and Environmental Epidemiology* 17 (2007), S. 549–558
- [US-EPA 1996] US-EPA: *Air Quality Criteria for Particulate Matter*. US–Environmental Protection Agency, Office of Research and Development, Washington, DC., April 1996 (EPA/600/P–95/001aF, Vol. I)
- [Van Roosbroeck u.a. 2007] VAN ROOSBROECK, S. ; JACOBS, J. ; JANSSEN, N.A.H. ; OLDENWENING, M. ; HOEK, G. ; BRUNEKREEF, B.: Long-term personal exposure to PM_{2.5}, soot and NO_x in children attending schools located near busy roads, a validation study. In: *Atmospheric Environment* 41 (2007), Nr. 16, S. 3381–3394
- [Van Roosbroeck u.a. 2008] VAN ROOSBROECK, S. ; LI, R. ; HOEK, G. ; LEBRET, E. ; BRUNEKREEF, B. ; SPIEGELMAN, D.: Traffic-Related Outdoor Air Pollution and Respiratory Symptoms in Children: The Impact of Adjustment for Exposure Measurement Error. In: *Epidemiology* 19 (2008), Nr. 3, S. 409–416
- [Wilson u. Spengler 1996] WILSON, R. ; SPENGLER, J.: *Particles in Our Air*. Harvard University Press, Boston, 1996. – ISBN 978–0–67424–077–3
- [Wright 1928] WRIGHT, P.G.: *The tariff on Animal and Vegetable Oils*. MacMillan, New York, 1928

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt, noch nicht einer anderen Prüfungsbehörde vorgelegt und noch nicht veröffentlicht habe.

Michael Sauer

München, 02. August 2011