Ludwig-Maximilians-Universität
Fakultät für Mathematik, Informatik und Statistik
Institut für Statistik

Treatment-Evaluationsproblematik

Seminararbeit zum Seminar
"Statistische Herausforderungen im Umgang
mit fehlenden bzw. fehlerbehafteten Daten"

von

Micha Fischer

15. März 2015

Betreuer: Prof. Dr. Thomas Augustin

Abstract

Diese Arbeit führt in ein spezielles Problem fehlender Daten, der Evaluation von Treatments bzw. Maßnahmen ein und gibt einen Überblick über gängige statistische Methoden zum Thema. Nach einer Beschreibung des „Rubin Causal Model“ und dessen Annahmen werden die zu schätzenden kausalen Parameter und deren Voraussetzungen vorgestellt. Im Folgenden wird speziell auf die Möglichkeiten der Schätzung kausaler Effekte im nicht-experimentellen Design, konkret auf Matching-Verfahren, Instrumentalvariablen, Regressionsansätze und Methoden für Paneldaten eingegangen. Das nachfolgende Kapitel beschreibt die Problemstellung aus Sicht partieller Identifikation. Zum Abschluss werden einige Methoden anhand einer Simulation verglichen.

Inhaltsverzeichnis

1 Einleitung

Die vorliegende Arbeit beschäftigt sich mit den statistischen Herausforderungen, die mit der Evaluation von Treatments (Behandlungen bzw. Maßnahmen) verbunden sind. Dabei können Behandlungen bzw. Maßnahmen medikamentöse Behandlungen, Fort- und Weiterbildungen, Einführungen von neuen Gesetzen und vieles mehr darstellen. Weil deren Durchführung oft teuer ist und die betroffenen Personen erheblich beeinflussen können, ist ein adäquates Evaluieren äußerst wichtig. Um den Erfolg bestimmen zu können, ist ein kausaler Schluss von Treatments auf Variablen, welche die Wirkung messen sollen (Outcomes), nötig. Beispielsweise kann die Wirkung eines blutdrucksenkenden Medikaments durch anschließende Messung des Blutdrucks erfolgen. Bei dieser naiven Herangehensweise ergeben sich allerdings Probleme:
Die potentielle Veränderung des Blutdrucks nach der Einnahme des Medikaments könnte auch auf andere Ursachen, z.B. die Umstellung der Ernährung zurückzuführen sein, wodurch die Schätzung verzerrt wäre, da der kausale Effekt nicht mehr isoliert betrachtet wird. Die Menge dieser potentiellen Confoundern ist unendlich, es ist somit nicht möglich alle zu messen und zu kontrollieren. Deshalb ist die Anwendung gängiger Methoden, wie einfache Regressionsmodelle nicht ausreichend um einen kausalen Zusammenhang unverzerrt schätzen zu können.
Im Folgenden wird das Fundamentalproblem kausaler Inferenz, eine zentrale Annahme und die Ziele der Analysen beschrieben. Danach werden experimentelles und nicht-experimentelles Design und die dazu gehörigen Methoden präsentiert. Wobei der Schwerpunkt auf Methoden bei nicht-experimentellen Situationen liegt. Dabei wird auch auf die Analyse kausaler Effekte mit Paneldaten eingegangen. Abschließend zeigt das Kapitel zur partiellen Identifikation noch eine andere Herangehensweise an die Problematik auf. Abschließend werden ausgewählte Methoden anhand einer Simulation verglichen.

1.1 Fundamentalproblem kausaler Inferenz

Der kausale Treatment-Effekt wird definiert als die Differenz des Outcomes Y bei erhaltenem Treatment ($D = 1$) und bei nichterhaltenem Treatment ($D = 0$). Allerdings müssen diese beiden Werte beim selben Subjekt i beobachtet werden können, um sicherzustellen, dass alle anderen Kovariablen gleich sind, die eine Auswirkung auf den Outcome Y haben, also eine „ceteris paribus-Situation“ vorhanden ist, wodurch alle Drittvariablen implizit kontrolliert werden können. Dadurch ergibt sich die folgende Formel für den Treatment-Effekt für das i-te Subjekt:

$$\delta_i = Y_i^1 - Y_i^0$$

Eine Beobachtung beider Zustände, Y_i^1 und Y_i^0, ist natürlich nicht möglich, da das i-te Subjekt entweder das Treatment erhalten hat oder nicht. Weil immer nur ein Outcome beobachtet werden kann, der andere aber nicht, wird dieser auch als kontrafaktisches Ereignis (counterfactual outcome) bezeichnet, da aber alle Realisationen von D und damit von Y möglich sind, heißen letztere auch potentielle Outcomes. Dieses Problem wird auch Fundamentalproblem kausaler Inferenz oder „Rubin Causal Model“ (Imbens u. Wooldridge, 2009, 4) genannt und kann als Problem fehlender Daten aufgefasst werden, da ausschließlich ein Ereignis und dessen Folgen für ein Subjekt, der Outcome, beobachtbar ist und der Rest fehlt. (Gangl 2010, 2-3; Caliendo u. Hujer 2006, 2-4; Legewie 2012, 127)
Das Konzept lässt sich natürlich auch auf Treatments mit mehr als zwei Ausprägungen (z.B. $D = 0$ “keine Therapie“, $D = 1$ „Therapie 1“, $D = 2$ „Therapie 2“, ...) übertragen. Der Einfachheit halber werden im Weiteren nur dichotome Treatmentvariablen betrachtet. Die folgenden Methoden stellen Versuche dar, das Fundamentalproblem kausaler Inferenz zu umgehen indem die fehlenden Werte durch sinnvolle, erhobene Werte ersetzt werden.

1.2 Stable Unit Treatment Value Assumption

Eine zentrale Annahme an das Modell ist, dass der Outcome Y_i^d des i-te Subjekts, welches das Treatment d erhalten hat, unabhängig von den Treatments und Outcomes aller anderen Subjekte ist. Diese „Stable Unit Treatment Value Assumption“ (SUTVA) ist im medizinischen Bereich, z.B bei Evaluation von neuen Medikamenten, meist erfüllt. Werden allerdings die Wirksamkeit von Impfungen gegen ansteckende Krankheiten untersucht, ist es sehr wohl möglich, dass die Häufigkeit der Impfung in der Bevölkerung, als auch deren Gesundheitszustand das Individuum und damit die ganze Evaluation beeinflussen.
Ist SUTVA verletzt, gibt es grundsätzlich zwei Lösungen: Zum einen, die Zusammenfassung der sich beeinflussenden Subjekte zu Clustern für die SUTVA erfüllt ist, oder eine Modellierung der Interaktionen. (Imbens u. Wooldridge, 2009, 9-10)
Die Annahme ist unter Fishers Null-Hypothese: $Y_i^d = Y_i^{d'}, \forall i, \forall d! = d'$, also der Hypothese, dass das Treatment keinerlei Effekt hat, immer erfüllt (Rubin, 1986, 2).

1.3 Ziele

Da eine Schätzung von $\delta_i = Y_i^1 - Y_i^0$, also die Schätzung eines Treatment-Effekts auf Individuenebene nicht möglich ist, wird versucht stattdessen den Erwartungswert (Average Treatment Effect, kurz ATE) $\mathbb{E}(Y^1 - Y^0) = \mathbb{E}(Y^1) - \mathbb{E}(Y^0)$ zu schätzen, der den Treatment-Effekt beschreibt, wenn die Zuteilung zum Treatment zufällig erfolgt (Caliendo u. Hujer, 2006, 3). Also eine „Missing Completely at Random“-Situation vorliegt. Definitionen und eine Einführung zu

den Missing-Mechanismen sind in den Vortragsmaterialien von Alexander Potatilo in diesem Seminar zu finden und werden darum hier als gegeben vorausgesetzt. Es wird in dieser Situation davon ausgegangen, dass die gegebenen Daten einer Zufallsstichprobe aus einer Grundgesamtheit entstammen, der Average Treatment Effect wird also auf Populationsebene geschätzt. (Imbens u. Wooldridge, 2009, 11)
Die Betrachtung der Subpopulation, die das Treatment erhalten hat, bietet eine weitere Möglichkeit der Schätzung eines kausalen Effekts. Dafür wird der sogenannte Average Treatment Effect on the Treated (ATT) definiert:

$$ATT = \mathbb{E}(Y^1 - Y^0|D = 1) = \mathbb{E}(Y^1|D = 1) - \mathbb{E}(Y^0|D = 1)$$

Dieser kommt hauptsächlich bei Beobachtungsstudien bzw. Befragungen zum Einsatz. Da dort davon auszugehen ist, dass die Zuteilung bzw. Teilnahme am Treatment nicht zufällig ist, kann es schwierig sein eine sinnvolle Vergleichsgruppe zu finden (Imbens u. Wooldridge, 2009, 11). Da $\mathbb{E}(Y^0|D = 1)$ ein kontrafaktischer Outcome ist, muss ein adäquater Ersatz gefunden werden. Die Verwendung vom beobachtbaren $\mathbb{E}(Y^0|D = 0)$ führt, da bei einer nicht-zufälligen Teilnahme am Treatment im Allgemeinen $\mathbb{E}(Y^0|D = 1) \neq \mathbb{E}(Y^0|D = 0)$ gilt, zu einer systematischen Verzerrung der Schätzung die „Selection Bias“ genannt wird. Dieser tritt auf, da sich Teilnehmer und Nicht-Teilnehmer systematisch durch beobachtbare und/oder nicht-beobachtbare Charakteristiken (Variablen) unterscheiden. Also die Outcome-Variable Y in den beiden Gruppen verschieden wären, selbst wenn das Treatment nicht verabreicht werden würde oder es keinerlei Effekt hätte. (Caliendo u. Hujer, 2006, 2-4)
Wenn dieses Confounding auf ungemessenen Variablen beruht, wird auch von unbeobachteter Heterogenität gesprochen (Fahrmeir u. a., 2009, 197).
Häufig ist in der Ökonometrie auch von Exogenität der Treatmentzuweisung die Rede. Diese liegt vor, wenn die Realisation der Treatmentvariable unabhänig von anderen Variablen ist, also kein Selection Bias vorliegt. Ist dies nicht erfüllt (Endogenität), kann durch berücksichtigen der confundierenden Variable wieder Exogenität erreicht werden. (Legewie, 2012, 128) Um das Ziel, eine unverzerrte Schätzung des Effekts eines Treatments auf einen Outcome zu erhalten, muss also zum einen gewährleistet sein, dass eine Zufallsstichprobe vorliegt und zum anderen, dass der Selection Bias eliminiert werden kann.

2 Experimentelles Design

Weil eine Messung des Effects δ_i für das i-te Subjekt nie möglich ist und auch ein j-tes Subjekt, dass sich nur durch die Ausprägungen in der Treatmentvariable D vom Subjekt i unterscheidet, nicht existiert, ist es nötig vergleichbare Gruppen von Subjekten zu betrachten, also Gruppen welche im Mittel die selben Ausprägungen an Drittvariablen haben.
Im experimentellen Design werden die Subjekte mittels Randomisierung zufällig auf (mindestens) zwei Gruppen aufgeteilt. Je größer die Gruppen werden, desto mehr nähern sich die Verteilungen der Kovariablen in den Gruppen einander an, da die Randomisierung einer Ziehung von (mindestens) zwei Stichproben aus der selben Verteilung entspricht. Die interessierende Outcome-Variable wird in allen Gruppen mindestens einmal vor und nach der Verabreichung von Treatment oder Placebo gemessen und kann dann verglichen werden. (Schuster 2009, 11-12; Steyer 1992, 3-5) Der Vergleich der mittleren Vorhermessungen zwischen den Gruppen (sowohl von Outcome-Variablen, als auch von weiteren erhobenen Variablen) kann zur Validierung des Randomisierungsprozesses genutzt werden. Bei einer potentiell guten Randomisierung sollten die Unterschiede zwischen den Gruppenmittelwerten möglichst gering sein. Allerdings sind kleine Gruppenabweichungen nur eine notwendige, aber keine hinreichende Bedingung für eine gute Randomisierung, da immer noch ungemessene Confounder existieren können.
Bei großen Gruppen sind allerdings Treatment und andere Covariablen nicht mehr von einander abhängig, wodurch der Selection Bias eliminiert werden kann. Nach einer Randomisierung ist es beispielsweise möglich, ein Regressionsmodell der Form $Y = \beta_0 + \beta_1 * D$ zu schätzen, was einem t-Test entspricht, oder die Differenz $\overline{Y}^1 - \overline{Y}^0$ zu bestimmen, beide Methoden ergeben erwartungstreue Schätzer für den Average Treatment Effect, weil die Treatmentvariable durch die Randomisierung exogen wird. (Imbens u. Wooldridge, 2009, 15)
Da in Situationen guter Randomisierung eine „Missing at Random“-Situation vorliegt, sind die für unverzerrte Schätzungen nötigen Methoden eher unkompliziert.
Das experimentelle Design stößt allerdings in Situationen an seine Grenzen, in denen eine Randomisierung nicht durchführbar oder aus ethischen Gründen nicht vertretbar ist. Beispielsweise werden in Beobachtungsstudien oder Umfragen die Teilnehmer oft retrospektiv, also nach Vorkommnissen, Handlungen oder Situationen in der Vergangenheit befragt, wodurch eine Randomisierung ausgeschlossen wird. Selbes gilt, wenn die erhobenen Daten erst später verwendet werden, also die Forschungsfrage zum Zeitpunkt der Erhebung noch gar nicht feststeht. Auch ist beispielsweise eine Randomisierung von unheilbar Kranken, von denen nur ein Teil ein potentiell wirksames Medikament erhält, aus ethischer Sicht bedenklich.

3 Nicht-experimentelles Design

Wann immer eine Randomisierung nicht möglich ist, muss auf Beobachtungsdaten zurückgegriffen werden. Wie oben beschrieben besteht die Hauptproblematik darin, dass die „Zuteilung“ zum Treatment nicht mehr unabhängig von Kovariablen oder der Outcome-Variable vor dem Treatment ist, also Endogenität vorliegt. Sind alle bedeutenden Variablen gemessen, liegt eine „Missing at Random“-Situation, andernfalls eine „Not Missing at Random“-Situation vor. Die folgenden Methoden zeigen, wie es möglich ist auch ohne Randomisierung eine valide Schätzung des Kausaleffektes zu erhalten oder zumindest den Selection Bias zu reduzieren.
Im Folgenden wird auf eine nachträgliche Bildung von Versuchs- und Kontrollgruppen eingegangen. Anschließend werden Regressionsansätze und die Verwendung von Instrumentalvariablen vorgestellt. Der Schluss dieses Kapitels befasst sich mit Methoden die bei der Analyse von Paneldaten Verwendung finden.

3.1 Matching-Verfahren

Die nun vorgestellten Verfahren dienen nicht dazu mehrere Datensätze zusammenzufügen, wie es im Vortrag von Katrin Hummrich zum Thema Statistical Matching der Fall war. Matching-Verfahren stellen den Versuch dar, in nicht-experimentellen Situationen, also bei Beobachtungsdaten nachträglich Versuchs- und Kontrollgruppen zu bilden, wie sie durch Randomisierung entstehen. Die Bildung der Gruppen erfolgt, indem jedem Subjekt aus der Treatmentgruppe ein Subjekt zugeordnet wird, dass kein Treatment erhalten hat, wobei darauf geachtet wird, dass die anderen gemessenen Kovariablen möglichst identische Ausprägungen haben (Caliendo u. Kopeinig, 2008, 2). Die Methode beruht auf der Annahme, dass eine „Missing at Random“-Situation gegeben ist. Das heißt, die Treatment-Variable ist unabhängig von den Outcomes, gegeben die gemessenen Kovariablen. Als Formel dargestellt:

$$D_i \perp (Y_i^0, Y_i^1)|\mathbf{X}_i$$

Wobei $\mathbf{X}_i$ ein Kovariablen-Vektor ist. Wird diese Annahme erfüllt, so unterscheiden sich die Verteilungen der Kovariablen von Treatment- und Kontrollgruppe nicht mehr voneinander. Die gewünschte „ceteris paribus-Situation", die auch durch Randomisierung erzeugt wird, ist gegeben. (Caliendo u. Hujer, 2006, 6) Da eine Unabhängigkeit bedingt auf gemessene Kovariablen angenommen wird, also dass keine weiteren ungemessenen Confounder existieren, kann es oft schwierig sein diese Annahme als erfüllt zu betrachten (Imbens u. Wooldridge, 2009, 21).
Damit das Verfahren funktioniert muss weiter angenommen werden, dass die Kovariablen keine perfekten Prediktoren für die Treatmentvariable sind, also dass $P(D = 1|\mathbf{X}) \in (0, 1)$ ist (Caliendo

u. Hujer, 2006, 6). Sonst ist es nicht gewährleistet, dass für eine oder mehrere Kombinationen von Ausprägungen von Kovariablen in der Treatmentgruppe ein passendes Subjekt gefunden werden kann, das kein Treatment erhalten hat.
Bei einem exakten Matching kann das Problem auftreten, dass keine adäquaten Partner für die Subjekte der Treatmentgruppe im Datensatz vorhanden sind, da um die ersten Annahme zu erfüllen möglichst viele der gemessenen Kovariablen verwendet werden, ergeben sich zu viele Kombinationsmöglichkeiten der Kovariablenausprägungen (Gangl u. DiPete, 2004, 17). Wenn beispielsweise ein Matching mit p dichotomen Variablen durchgeführt werden soll, so ergeben sich 2^p Möglichkeiten diese Merkmale zu kombinieren. Schon in diesem einfachen Beispiel verringert sich die Wahrscheinlichkeit vergleichbare Merkmalsträger zu finden mit zunehmender Variablenzahl schnell.
Aus diesem Grund wird oft auf sogenannte Balancing-Scores zurückgegriffen, die dieses Problem beheben sollen. Am häufigsten angewendet wird der Propensity-Score, welcher die Möglichkeit des Matchings von Subjekten unter der Verwendung von fast beliebig vielen Kovariablen bietet, indem die Dimensionen der zu matchenden Variablen auf eine einzige reduziert werden. Durch diese Aggregation entsteht eine vollkommene Ordnung, welche eine einfache Zuordnung von vergleichbaren Merkmalsträgern ermöglicht. Der entstehende Wert kann als die Wahrscheinlichkeit für die i-te Person das Treatment D zu erhalten, gegeben einen Vektor an Kovariablen $\mathbf{X}$ beschrieben werden :

$$e(x_i) = P(D_i = 1|\mathbf{X}_i = \mathbf{x}_i)$$

Wenn der einfachste Fall einer dichotomen Treatment-Variable vorliegt, kann der Propensity Score durch ein logistisches Regressionsmodell für alle Merkmalsträger geschätzt werden (vgl. Fahrmeir u. a. 2009, 189-197). Die Qualität des Verfahrens ist natürlich stark von der Güte der gegebenen Kovariablen abhängig, also wie viel Einfluss diese auf die Treatment-Variable haben. Nach der Berechnung können den Subjekten der Treatmentgruppe die Subjekte mit dem ähnlichsten Propensity-Score mittels eines Matching-Algorithmus zugeordnet werden. (D ´Agostino, Ralph B. Jr., 1998, 2266-2269)
Da immer davon ausgegangen wird, dass alle wichtigen Kovariablen berücksichtigt wurden, dies aber nicht unbedingt der Fall sein muss, kann durch Matchingverfahren nicht sichergestellt werden, dass der Selection Bias vollkommen verschwindet. Da aber zumindest eine Bias-Reduktion vorgenommen werden kann, ist der Einsatz von Matchingverfahren sicherlich sinnvoller als ein naiver Vergeich der Gruppen.

3.2 Regressionsansatz

Eine Schätzung kausaler Effekte ist auch mittels Regressionsmodellen durchführbar. Hierbei können seperate Modelle für die Gruppe der Behandelten, $\mu_1(x) = \mathbb{E}(Y_i|D = 1, X_i = x)$, und der Nicht-Behandelten, $\mu_0(x) = \mathbb{E}(Y_i|D = 0, X_i = x)$, mit der selben Kovariablenstruktur geschätzt und anschließend verglichen werden. Eine Schätzung des Kausaleffekts kann beispielsweise durch $\widehat{ATE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))$ erhalten werden. (Imbens u. Wooldridge, 2009, 23-24)
Es wird angenommen, dass die Fehlerterme unabhängig vom Treatment gegeben die Kovariablen sind. Dem aufmerksamen Leser fällt sicherlich die ähnlich zu den oben beschriebenen Matchingverfahren auf, die auf der Annahme der Unabhängigkeit von Outcome und Treatment gegeben die Kovariablen beruhen. Tatsächlich lässt sich zeigen, dass sich die beiden Verfahren annähern, wenn im Regressionsansatz die Kovariablen nicht nur einfach, sondern zusätzlich in höheren Potenzen und Interaktionen aufgenommen werden. (Caliendo u. Hujer, 2006, 7) Der Unterschied besteht insofern, dass einfaches Matching nonparametrisch und der Regressionsansatz parametrisch ist. Eine weitere Regressionsbassierte Methode wird im Kapitel 3.4 behandelt.

3.3 Instrumentalvariablen

Eine Möglichkeit mit unbeobachteter Heterogenität umzugehen bieten Instrumentalvariablen. Um als Instrumentalvariable Z dienen zu können, muss die Variable zum einen mit der Treatmentvariable in Zusammenhang stehen, darf zum anderen aber nicht mit der Outcome-Variable zusammenhängen. Der Outcome wird von Z also nur indirekt über die Treatment-Variable beeinflusst. Erfüllt eine Variable dieses Kriterium, so kann sie zur Schätzung des kausalen Effekts eingesetzt werden. (Caliendo u. Hujer, 2006, 8)
Zur Schätzung des kausalen Effekts wird nur der durch Z erklärte Anteil der Treatmentvariable verwendet wodurch kein Selection Bias mehr vorhanden ist (Legewie, 2012, 15). Wenn beispielsweise die Auswirkungen einer Gesetzesänderung in einem Land untersucht werden soll, so können die Outcome-Variablen von diesem und einem anderen Land, welches noch die alte Gesetzeslage hat, verglichen werden. Wenn die Outcome-Variablen vorher auf dem selben Niveau lagen, so ist die „Land-Variable“ ein Instrument. In diesem extremen Beispiel wird das Treatment komplett durch das Instrument erklärt, es liegt ein natürliches (bzw. Quasi-) Experiment vor, wodurch auch bei ungemessenen Confoundern eine unverzerrte Schätzung des kausalen Effekts möglich wird (Gangl, 2010, 36).
Leider ist das Auffinden einer solchen Variablen in vielen Situationen nicht möglich (Caliendo u. Hujer, 2006, 8), darum kann diese Methode auch nur eingeschränkt eingesetzt werden.

3.4 Treatmentevaluation mit Paneldaten

Die Struktur von Paneldaten, die Messung zu mehreren Zeitpunkten am selben Subjekt, bietet eine gute Möglichkeit kausale Effekte zu schätzen, da eine Betrachtung individueller Outcome-Verläufe vor und nach dem Treatment möglich ist. Des Weiteren kann die Wirkung des Treatments im Zeitverlauf untersucht werden, es ist beispielsweise möglich, dass eine Wirkung erst später eintritt oder diese nur für kurze Zeit anhält. Diese Fragestellung kann mittels Paneldaten untersucht werden. Auch in diesem Kontext kann eine Anwendung der oben beschriebenen Methoden wie Matching-Verfahren möglich sein. Der folgende Differences-in-Differences-Schätzer kann auch ohne Paneldaten verwendet werden, solange mindestens eine Messung vor dem Treatment und eine danach erfolgt. Die anschließend vorgestellten Fixed-Effects-Modelle sind allerdings auf eine Panelstruktur angewiesen. Beide Verfahren haben den Vorteil, dass trotz unbeobachteter Heterogenität, also unbeobachteten Confounder-Variablen, eine unverzerrte Schätzung möglich ist. (vgl Brüderl (2010))

3.4.1 Differences-in-Differences-Schätzer

Wie der Name schon vermuten lässt, werden beim Differences-in-Differences-Schätzer (DID) Differenzen betrachtet, also ein relativer Wert des kausalen Effekts geschätzt, indem die mittleren Veränderungen der Outcome-Variable in Treatment- und Kontrollgruppe verglichen werden:

$$DID = (\bar{Y}_t^1 - \bar{Y}_{t'}^0 | D = 1) - (\bar{Y}_t^0 - \bar{Y}_{t'}^0 | D = 0)$$

Wobei $Y_{t'}$ die Outcome-Variable vor dem Treatment und Y_t nach dem Treatment bezeichnen. Es sind also sowohl Vorher- als auch Nachhermessungen in Kontroll- und Vergleichsgruppe nötig, um den Schätzer bestimmen zu können, wobei die Mittelung bewirkt, dass die Messungen nicht unbedingt an den selben Subjekten vorgenommen werden muss (Brüderl u. Ludwig, 2015, 11). Dabei wird angenommen, dass $\mathbb{E}(Y_t^0 - Y_{t'}^0 | D = 1) = \mathbb{E}(Y_t^0 - Y_{t'}^0 | D = 0)$ gilt, also der Unterschied der Differenzen nur durch das Treatment entsteht bzw. der Selection Bias nur von Variablen verursacht wird, die konstant über die Zeit wirken. Durch die Differenzenbildung können solche und auch lineare Effekte berücksichtigt werden. (Caliendo u. Hujer, 2006, 9-10) Die Anwendung des DID-Schätzers ist weit verbreitet und kann auch gut mit anderen Methoden, wie beispielsweise Matching kombiniert werden.(Gangl, 2010, 34)

3.4.2 Fixed-Effects-Regression

Wenn Paneldaten, also mehrere Messungen zu verschiedenen Zeitpunkten am selben Subjekt vorliegen, ist es möglich, individuelle Differenzen zwischen den Messzeitpunkten zu betrachten, anstatt Gruppendifferenzen zu berechnen wie es beim DID-Schätzer der Fall ist. Dies ist die Grundidee der Fixed-Effects-Regression. Um nur diese individuellen Differenzen über die Zeit in den Daten zu betrachten gibt es (unter anderem) zwei Möglichkeiten:
Zum einen kann in einem (linearen) Regressionsmodell für jedes Subjekt ein eigener Intercept geschätzt werden, wobei alle $Subjekte * Messungen$ Beobachtungen als unabhängig angenommen werden. (Brüderl u. Ludwig, 2015, 10)
Die Modellformel für ein Subjekt i zu einem Zeitpunkt t ergibt sich dann wie folgt:

$$y_{it} = \alpha_i + \mathbf{x}_{it}\beta + \epsilon_{it} \tag{3.1}$$

Der individuelle Intercept α_i kann auch als subjektspezifische Fehlerkomponente, welche die unbeobachtete und zeitkonstante Heterogenität des Individuums enthält, aufgefasst werden und darf auch mit den Kovariablen $\mathbf{x}_i$ korreliert sein. (Brüderl u. Ludwig, 2015, 3)
Die Verwendung von individuenspezifischen Intercepts ist nicht in allen Situationen möglich. Weil immer $N-1$ Parameter zusätzlich zu den Koeffizienten der Kovariablen geschätzt werden müssen ist dies mit Standard-Software-Paketen für große N schnell nicht mehr möglich.
Ein Ausweg bietet eine Transformation der Daten, sodass nur noch die Variation innerhalb der Subjekte über die Zeit in den Daten vorhanden ist. Diese Transformation entspricht einer Mittelwertsbereinigung und ist auch als Within-Transformation bekannt. Um die Daten zu transformieren ist es zunächst nötig, die Komponenten aus Gleichung (3.1) über alle Zeitpunkte t für alle Subjekte zu mitteln wodurch sich

$$\overline{y}_i = \alpha_i + \overline{\mathbf{x}}_i\beta + \overline{\epsilon}_i \tag{3.2}$$

mit $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\overline{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}$ und $\overline{\epsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \epsilon_{it}$ ergibt. Abschließend kann (3.2) von (3.1) subtrahiert werden, um die folgende Gleichung zu erhalten:

$$y_{it} - \overline{y}_i = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)\beta + (\epsilon_{it} - \overline{\epsilon}_i) \tag{3.3}$$

Nach der Transformation der Daten können auch diese in einem (linearen) Modell verwendet werden um kausale Effekte zu schätzen. Da, wie in Gleichung (3.3) ersichtlich, der subjektspezifische Intercept nicht mehr existiert, muss dieser auch nicht mehr in die Modellformel aufgenommen werden und die problematische Berechnung entfällt (Brüderl u. Ludwig, 2015, 3-4). Falls subjektspezifische Verläufe über die Zeit in der Fragestellung von Belang sind, ist die Verwendung von

individuellen Intercept-Termen natürlich obligatorisch.
Da gezeigt werden kann, dass die Fehler eines Subjektes i negativ korreliert sind, also $Cov(\epsilon_{it}, \epsilon_{is}) = 0$ verletzt ist, kann hier der Standardfehler nicht auf die klassische Weise geschätzt werden. Sogenannte „Panel-robuste Standardfehler“ versuchen diese Korrelationen zu berücksichtigen. Es stellt sich allerdings heraus, dass in Datensituationen mit vielen Merkmalsträgern und wenigen Messzeitpunkten diese Standardfehler zu klein geschätzt werden. Eine unverzerrte Schätzung dieser Standardfehler ist leider nicht in allen Situationen möglich und die daraus folgenden Ergebnisse sollten darum immer kritisch betrachtet werden. (Brüderl u. Ludwig, 2015, 10)
Fixed-Effects-Modelle können auch als gemischte Modelle angesehen werden in denen keine zufälligen Effekte vorkommen, sie sind also ein Spezialfall. Wie auch im gemischten Modell ist es hier möglich, neben subjektspezifischen Intercepts noch individuelle Steigungen zu modellieren wenn die Subjekte hier stark variieren. Dieses Vorgehen wird allerdings noch nicht häufig angewendet. (Brüderl u. Ludwig, 2015, 12-15)
Da eine, durch die Transformation der Daten bzw. der Schätzung der subjektbezogenen Intercepts induzierte Varianzreduktion vorgenommen wird, nämlich genau die Varianz zwischen den verschiedenen Subjekten (auch Between-Varanz genannt) eliminiert wird, also Information verloren geht, ist die Schätzung des Standardfehlers in FE-Modellen oft größer als beispielsweise in gemischten Modellen. Die FE-Schätzung ist allerdings auch bei ungemessenen zeitkonstanten Confoundern unverzerrt. (Brüderl, 2010, 14-15)

4 Partielle Identifikation

Da alle Modelle auf Annahmen beruhen, können sich Situationen ergeben in denen das Resultat einer Analyse genauer erscheint als es eigentlich ist, da getroffene Annahmen nicht oder nur teilweise erfüllt sind. Dieser Sachverhalt führt zum Gesetz der abnehmenden Glaubwürdigkeit, dass besagt, die Glaubwürdigkeit von Ergebnissen verringert sich je stärker die benötigten Annahmen sind. (Manski, 2003, 2) Das Konzept der Partiellen Identifikation verzichtet nun auf eine oder mehrere Annahmen die ein Modell eindeutig identifizieren. Dadurch sind die durch Analysen erhaltenen Ergebnisse im Allgemeinen nicht mehr eindeutig. Je nach Hinzu- oder Wegnahme von Annahmen kann der Bereich, die Identifikationsregion, verkleinert oder vergrößert werden.
In diesem Kapitel wird anfangs die Logik partieller Identifikation im Kontext fehlender Daten eingeführt und darauf aufbauend für die Treatmentevaluation beschrieben.

4.1 Fehlende Daten

Gegeben sei eine Grundgesamtheit N, wobei das i-te Subjekt daraus einen Outcome y_i aus Y besitzt, ähnlich zu den oberen Kapiteln. Neu ist, dass eine messbare Funktion $y : N \to Y$ definiert werden kann. Da die Grundgesamtheit ein Wahrscheinlichkeitsraum (N, Ω, P) ist, ist y eine Zufallsvariable welche die Verteilung $P(y)$ besitzt. Ziel ist es $P(y)$ durch eine Zufallsstichprobe aus N zu schätzen und dabei können Werte aus der Stichprobe nicht beobachtbar sein. Die Variable z dient als Indikator für fehlende Werte, wobei $z = 1$ bedeutet, dass y beobachtet werden konnte. Die Verteilung $P(y)$ kann mit dem Satz der totalen Wahrscheinlichkeit folgendermaßen beschrieben werden:

$$P(y) = P(y|z = 1)P(z = 1) + P(y|z = 0)P(z = 0)$$

Durch die Ziehung einer Zufallsstichprobe können Schätzer für $P(y|z = 1)$ und $P(z)$ erhalten werden. Die Verteilung der fehlenden y-Werte $P(y|z = 0)$ ist natürlich nicht schätzbar. Eine Identifikationsregion $H[P(y)]$ kann in dieser Situation folgendermaßen angegeben werden:

$$H[P(y)] = [P(y|z = 1)P(z = 1) + \gamma P(z = 0), \gamma \in \Gamma_Y]$$

Γ_Y steht hier für den Raum aller Wahrscheinlichkeitsmaße auf Y. Da die Identifikationsregion eine gewichtete Summe aus den $P(y|z = 1)$ und den γ aus Γ_y ist, hängt ihre Größe von den Gewichten $P(z = 0)$ und $P(z = 1)$, also von dem Anteil an fehlenden Daten, ab. Sie ist eine einelementige Menge für $P(z = 0) = 0$ und damit vollständig identifiziert. Für $0 < P(z = 0) < 1$ ergibt sich eine partielle Identifikation, für $P(z = 0) = 1$ ist die Verteilung überhaupt nicht identifiziert. Durch Annahmen ist es möglich Identifikationsregionen weiter zu verkleinern. Beispielsweise könnte die Annahme $P(y|z = 1) = P(y|z = 0)$, also wenn eine „Missing Completely at Random“-Situation vorausgesetzt wird, $P(y)$ wieder komplett identifizieren. (Manski, 2003, 6-7, 26-27)
Auch für bedingte Verteilungen $P(y|X = x)$ mit gegebenen und vollständig beobachteten Kovariablen X können Identifikationsregionen angegeben werden:

$$H[P(y|X = x)] = [P(y|X = x, z_y = 1)P(z_y = 1|X = x) + \gamma P(z_y = 0|X = x), \gamma \in \Gamma_Y]$$

Für diese gelten die obigen Zusammenhänge analog. (Manski, 2003, 41)

4.2 Treatmentevaluation

Im Kontext des Fundamentalproblems kausaler Inferenz, die als Fehlende-Daten-Situation aufgefasst werden kann, ergeben sich ähnliche Konstrukte, allerdings müssen hier noch weitere Defini-

tionen vorgenommen werden: Die Variable $z_i \in D$ gibt an, welches Treatment dem i-ten Subjekt zugeordnet wird und $y_i = y_i(z_i)$ beschreibt den zugehörigen Outcome, wobei $y_i(d)$ für $d \neq z_i$ ein kontrafaktisches Ereignis ist. Außerdem ist $z : N \to D$ die Funktion, welche den Subjekten aus N die Treatments zuordnet, die sie tatsächlich erhalten. Es ergibt sich die Verteilung $P(y(d)|X = x)$ für gegebenen Kovariablen X wenn alle Subjekte mit den Kovariablen x das Treatment d erhalten. Da aber die Subjekte für gegebene Kovariablen verschiedene Treatments erhalten, wird daraus eine Menge an Verteilungen $\{P(y(d)|X), d \in D\}$. Das Ziel, welches auch Selektionsproblem genannt wird, ist die Identifikation dieser Verteilungen durch Annahmen und einer Stichprobe, welche Schätzungen für $P(y, z|x)$ und die Verteilungen der Kovariablen $P(x)$ liefert. Durch die Betrachtung der Formel für $P(y(d)|X = x)$ werden die Parallelen zur Fehlenden-Daten-Situation noch deutlicher:

$$\begin{aligned} P(y(d)|X = x) &= P(y(d)|X = x, z = d)P(z = d|X = x) + P(y(d)|X = x, z \neq d)P(z \neq d) \\ &= P(y|X = x, z = d)P(z = d|X = x) + P(y(d)|X = x, z \neq d)P(z \neq d) \end{aligned}$$

Die dazugehörige Identifiaktionsregion ergibt sich mit einer Stichprobe folgendermaßen:

$$H[P(y(d)|X = x)] = [P(y|X = x, z = d)P(z = d|X = x) + \gamma P(z \neq d|X = x), \gamma \in \Gamma_Y]$$

Da es nicht möglich ist durch eine Stichprobe Informationen über $P(y(d)|X = x, z \neq d)$ zu erhalten, kann diese Region nicht vollständig identifiziert werden, wenn nur die auf Basis der Stichprobe gewonnenen Informationen verwendet werden. (Manski, 2003, 99-101)
Im Falle einer Randomisierung wie sie in Kapitel 2 vorgestellt wird, ist die vollständige Identifikation möglich, da durch die Randomisierung eine Unabhängigkeit des gesuchten Verteilung $P(y(d)|X = x)$ von der Funktion z stattfindet, und damit

$$P(y(d)|X = x) = P(y(d)|X = x, z = d) = P(y|X = x, z = d), \forall d \in D, \forall x \in X$$

gilt. In Situationen ohne Randomisierung ist die Gültigkeit dieser Annahme wie in Teil 3 dargestellt, fraglich. (Manski, 2003, 101-102)
Somit sind dort andere Annahmen nötig um eine vollständige Identifkation zu erreichen.

5 Simulation

Das Ziel dieser Simulation ist der Vergleich von einigen vorgestellten Methoden im nicht-experimentellen Design anhand einer künstlichen Datensituation in welcher die Verteilungen der Daten und die Zusammenhänge zwischen den einzelnen Variablen bekannt sind. Im Folgenden wird auf die erstellte

Datengrundlage eingegangen, danach die Anpassung der verwendeten Methoden an die konkrete Situation beschrieben und abschließend die Ergebnisse anhand einer Graphik verglichen.

5.1 Datenerzeugung

Das Paket „mvtnorm“ welches in der Software R enthalten ist, ermöglicht es ohne großen Aufwand Zufallszahlen aus multivariaten Normalverteilungen zu ziehen. Diese wurde genutzt um Beobachtungen mit den in Tabelle 1 gezeigten Variablen zu erzeugen.

Variable	Beschreibung
D	Treatmentvariable, dichotom
X_1	Kovariable, normalverteilt
X_2	Kovariable, dichotom, zeitkonstant
Y	Outcomevariable, normalverteilt
t	Indikator für die beiden Messzeitpunkte
$treat$	Indikator für Treatmentgruppe
ID	Subjekt-ID

Tabelle 1: Übersicht der erzeugten Variablen

Jedes Subjekt wurde hierbei zu zwei Zeitpunkten beobachtet, einmal vor dem Treatment und einmal danach, was einer Paneldatenstruktur entspricht. Des Weiteren wurden die in Bild 5.1 beschriebene Abhängigkeitsstruktur mittels Korrelation von Y und X_1, unterschiedliche Mittelwerte von Y für die beiden Ausprägungen von $treat$ und X_2 und unterschiedliche Gruppenstärken ($treat = 0, X_2 = 0$: 500; $treat = 1, X_2 = 0$: 250; $treat = 0, X_2 = 1$: 300; $treat = 1, X_2 = 1$: 200) für den Zeitpunkt $t = 0$ erzeugt.

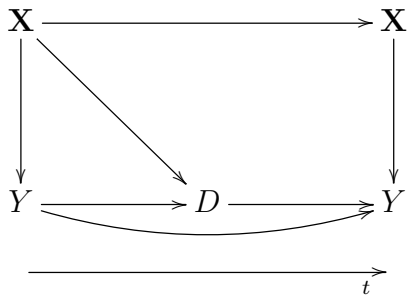

Abbildung 5.1: Zusammenhang zwischen den Variablen

Anschließend wurden für alle $n = 1250$ Subjekte zum Zeitpunkt $t = 1$ Beobachtungen generiert, wobei dazu die Variable X_1 aus einer Normalverteilung mit dem Wert von X_1 bei $t = 0$ als Erwartungswert verwendet wurde. Der Wert der Y-Variable ergibt sich für jedes Subjekt durch

Ziehungen aus einer Normalverteilung mit dem Erwartungswert, der sich zusammensetzt aus der Summe von: Dem Wert von Y bei $t = 0$, dem Treatmenteffekt multipliziert mit D und dem Effekt von X_1 zum Zeitpunkt $t = 1$ multipliziert mit dem Wert von X_1 bei $t = 1$. Da nur für die Treatment-Gruppe $D = 1$ gilt, erhält auch nur diese den Treatmenteffekt. Für die ganze Simulation wurde der Treatmenteffekt$= 10$ und der Effekt von X_1 auf $-0,7$ gesetzt.

5.2 Verwendete Methoden

Auf die Daten wurden die Methoden, die im Kapitel 3 beschrieben und für die Situation geeignet sind, angewandt. Falls möglich, wurde einmal eine „Missing at Random"- Situation, indem alle verfügbaren Kovariablen verwendet wurden, und einmal eine „Not Missing at Random"-Situation, bei der die Kovariable X_2 nicht zur Schätzung benutzt wurde, dargestellt.

5.2.1 Einfacher Mittelwertsvergleich

Zur Veranschaulichung wird die Differenz der Mittelwerte der Outcome-Variable Y von Treatment- und Kontrollgruppe zum Zeitpunkt $t = 1$ betrachtet, wie sie unter einer „Missing Completely at Random"-Situation zu einer unverzerrten Schätzung führen würde. Dadurch kann gezeigt werden, wie effektiv die einzelnen Methoden in der Reduktion des Bias sind.

5.2.2 Differences-in-Differences-Schätzer

Aufbauend auf dem Mittelwertsvergleich werden hier auch die unterschiedlichen Gruppen-Niveaus vor dem Treatment in Y, also zum Zeitpunkt $t = 0$ in der Analyse betrachtet und somit versucht den Selection-Bias zu reduziert.

5.2.3 Propensity-Score-Matching

Mit dem R-Paket „MatchIt" ist eine Berechnung des Propensity-Scores, ein anschließendes Matching der Subjekte und die Erstellung eines Datensatzes mit den „gematchten" Subjekten durchführbar. Hier kann zum ersten Mal zwischen „MAR" und „NMAR" unterschieden werden. Im ersten Fall ergibt sich die folgende Modellformel:

$$treat = Y + X_1 + X_2$$

Es wird die Variable X_2 als gegeben betrachtet und das logistische Regressionsmodell mit den Daten zum Zeitpunkt $t = 0$ geschätzt und daraufhin das Matching vorgenommen. Zur Evaluation des Vorgangs ist es sinnvoll die Mittelwerte der verwendeten Kovariablen für die beiden Gruppen

und beiden Datensätze (vor dem Matching und nach dem Matching) zu betrachten. Wie zu erwarten unterscheiden sich diese nach dem Matching weniger voneinander.
Für die Matching-Daten folgt nun ein einfacher Mittelwertsvergleich wie unter 5.2.1 beschrieben. Außerdem wurde der Differences-in-Differences-Schätzers angewendet um die Bias-Reduktion noch weiter zu verbessern.
Das Vorgehen in der „Not Missing at Random"-Situation unterscheidet sich nur geringfügig. Hier wird diese Modellformel verwendet:

$$treat = Y + X_1$$

Also die Variable X_2 nicht zur Schätzung benutzt, Mittelwertsvergleich und Differences-in-Differences-Schätzer werden analog angewendet.

5.2.4 Regressionsansatz

Es erfolgt eine Regression mit der Outcomevariable Y zum Zeitpunkt nach der Treatmentanwendung als abhängige Variable zum einen für die Kontrollgruppe und zum anderen für die Treatmentgruppe, wobei die Modellformel gleich bleibt und wie folgt lautet:

$$Y_{t=1} = X_{1,t=0} + X_2 + X_{1,t=1}$$

Es ist, da X_2 im Modell verwendet wird, eine „Missing at Random"-Situation, für die „Not Missing at Random"-Situation fehlt diese Variable in der Modellformel. Die Schätzung für den kausalen Effekt ist anschließend durch die Bildung des arithmetischen Mittels aus den Diffrenzen der Vorhersagen durch die beiden Modelle, wie in Kapitel 3.2 beschrieben, berechnet worden.

5.2.5 Fixed-Effects-Regression

Die Modellierung erfolgt hier nach der folgenden Formel:

$$Y = D + t + X_1 + ID$$

Wobei die Variablen D, t und ID als Dummy-Variablen aufgenommen werden, durch die Dummy-Codierung der Subjekt-ID ergeben sich subjektspezifischen Intercept-Terme im Modell. Um die Auswirkungen verschiedener Modellformeln darzustellen wurde zusätzlich die folgende Formel verwendet:

$$Y = D + t * X_1 + ID$$

Neben den Haupteffekten ist auch die Interaktion zwischen der Zeit t und der Variable X_1 im Modell enthalten. Die Schätzung für den kausalen Effekt ergibt sich durch die Schätzung des Regressionsparameters der Variablen D. Eine Unterscheidung zwischen „NMAR“ und „MAR“ ist hier nicht sinnvoll, da X_2 eine zeitkonstante Variable ist.

5.3 Ergebnisse

Nachdem die Schätzungen der kausalen Effekte durchgeführt wurden, können diese verglichen werden. Die folgende Graphik 5.2 zeigt die erhaltenen Punktschätzer.

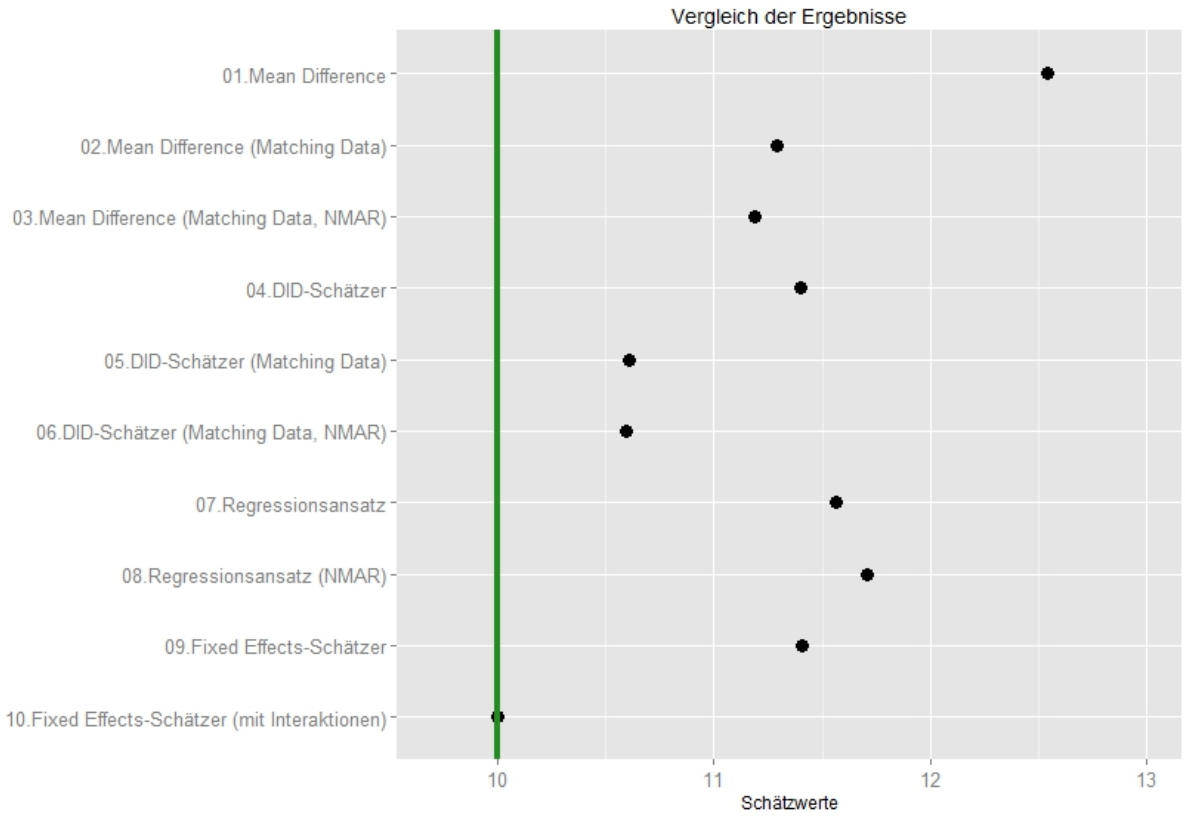

Abbildung 5.2: Gegenüberstellung der Ergebnisse

Auf der X-Achse sind die erhaltenen Werte auf einer Skala von 9,5 bis 13 abgetragen. Die Y-Achse dient nur dazu die einzelnen Werte den entsprechenden Methoden zuzuordnen. Die vertikale grüne Linie bei $X = 10$ zeigt den wahren Wert zugrundeliegenden kausalen Effekt von 10 an. Je kleiner der Abstand zwischen Punkt und Linie, desto unverzerrter/besser ist die Schätzung.
Wie erwartet ist die Verzerrung bei einem einfachen Mittelwertsvergleich (01. Mean Difference) mit etwa 2,5 am größten. Durch die Verwendung des Matchings lässt sich der Bias reduzieren, dies zeigt sich sowohl beim einfachen Mittelwertsvergleich mit den Matching-Daten (02. Mean Diffe-

rence (Matching Data), der Abstand zum wahren Wert beträgt hier nur noch etwa 1,25, als auch bei der Verwendung des Differences-in-Differences-Schätzers (Vergleich von 04. DID-Schätzer und 05. DID-Schätzer (Matching Data)), bei dem der Abstand ca. 0,6 ist. Auch durch den Regressionsansatz kann der Bias auf 1,55 bzw. 1,7 vermindert werden (07.Regressionsansatz, 08.Regressionsansatz (NMAR)). Die Schätzung durch die Fixed-Effects-Regression ohne Interaktionen reduziert den Bias ähnlich wie die obigen Verfahren auf etwa 1,4. Im Modell, dass Interaktionen enthält kann der Bias vollkommen reduziert werden. Dies zeigt, dass auch hier eine falsche Modellierung zu einer Verzerrung führen kann und dieser viel Aufmerksamkeit geschenkt werden sollte. Die Unterschiede zwischen „NMAR“ und „MAR“ sind in diesem Anwendungsbeispiel eher gering.
Auf Grund der Einfachheit dieser Simulation sollten die Ergebnisse nicht überinterpretiert werden. In anderen Datensituationen ist möglicherweise eine andere Methode effektiver und manche Konzepte sind vielleicht gar nicht anwendbar, wenn beispielsweise die Panel-Daten-Struktur nicht gegeben ist. Auch wenn viele der benutzten Methoden noch weiter verfeinert und besser an die gegebenen Daten angepasst werden können, zeigt sich, dass selbst mit einer simplen Verwendung der Methoden eine Reduktion des Selection-Bias möglich ist. Darum sollten diese Methoden immer zur Anwendung kommen, um eine möglichst gute Schätzung eines kausalen Effekts zu erzielen.

6 Schluss

Um eine Schätzung von kausalen Effekten durchzuführen und damit eine Maßnahme evaluieren zu können sind je nach Situation verschiedene Strategien anwendbar. Die anfangs erwähnten Kriterien für unverzerrte Schätzungen in diesem Kontext werden meist nicht durch eine alleinige Methode abgedeckt. Experimentelle Designs besitzen meiste hohe interne Validität, sie bieten die beste Möglichkeit zur Drittvariablenkontrolle und damit zur Eliminierung des Selection Bias, aber wenig externe Validität, eine Zufallsstichprobe ist oft nicht gegeben, was die Generalisierbarkeit der Ergebnisse erschwert. Im nicht-experimentellen Design verhält es sich meist genau anders herum. (Diekmann, 2004, 344-345)
Um ein glaubwürdiges Ergebnis zu erhalten ist eine Kombination von verschiedenen Strategien und der kritische Umgang mit den Modellannahmen zwingend. Die Glaubwürdigkeit der Ergebnisse hängt von den Annahmen der verwendeten Methoden und insbesondere vom zugrundeliegenden Missing-Mechanismus ab. Sind einfache Methoden in „Missing Completely at Random“-Fällen noch anwendbar, so wird bei „Missing Completely at Random“- oder „Not Missing at Random“-Situationen eine komplexere Methodik, mit oftmals starken Annahmen, benötigt um unverzerrte Ergebnisse zu erzielen. Da es außerhalb des experimentellen Designs nicht möglich ist den Missing-Mechanismus exakt zu identifizieren und damit die Annahme eines bestimmten Mecha-

nismus immer unerfüllt sein kann, so ist die Schätzung des Kausaleffekts immer mit Skepsis zu betrachten. In Anbetracht dessen ist es umso wichtiger, dass sich nicht auf eine einzelne Methode verlassen wird, sondern Methoden aus experimentellen und nicht-experimentellen Design kombiniert werden, um eine größtmögliche Sicherheit zu erzielen.

Literatur

[Best u. Wolf 2015] BEST, Henning (Hrsg.) ; WOLF, Christof (Hrsg.): *The Sage handbook of regression analysis and causal inference*. 1. publ. Los Angeles Calif. u.a : SAGE, 2015 (SAGE reference). – ISBN 978–1–4462–5244–4

[Brüderl 2010] BRÜDERL, Josef: Kausalanalyse mit Paneldaten. In: WOLF, Christof (Hrsg.): *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden : VS Verl. für Sozialwissenschaften, 2010. – ISBN 978–3531163390, S. 963–994

[Brüderl u. Ludwig 2015] BRÜDERL, Josef ; LUDWIG, Volker: Fixed-Effects Panel Regression. In: BEST, Henning (Hrsg.) ; WOLF, Christof (Hrsg.): *The Sage handbook of regression analysis and causal inference*. Los Angeles Calif. u.a : SAGE, 2015 (SAGE reference). – ISBN 978–1–4462–5244–4

[Caliendo u. Hujer 2006] CALIENDO, Marco ; HUJER, Reinhard: The microeconometric estimation of treatment effects: An overview. In: *Allgemeines statistisches Archiv : AStA ; journal of the German Statistical Society* 90 (2006), Nr. 1, S. 199–215

[Caliendo u. Kopeinig 2008] CALIENDO, Marco ; KOPEINIG, Sabine: Some practical guidance for the implementation of propensity score matching. In: *Journal of economic surveys* 22 (2008), Nr. 1, S. 31–72

[D ´Agostino, Ralph B. Jr. 1998] D ´AGOSTINO, RALPH B. JR.: Propensity Score Methods for Bias Reduction in the Comparison of a Treatment to a Non-Randomized Control Group. In: *Statistics in Medicine* (1998), S. 2265–2281

[Diekmann 2004] DIEKMANN, Andreas (Hrsg.): *Methoden der Sozialforschung*. 2004

[Fahrmeir u. a. 2009] FAHRMEIR, Ludwig ; KNEIB, Thomas ; LANG, Stefan: *Regression: Modelle Methoden und Anwendungen*. Berlin [u.a.] : Springer, 2009 (Statistik und ihre Anwendungen). http://dx.doi.org/10.1007/978-3-642-01837-4. – ISBN 978–3–642–01837–4

[Gangl 2010] GANGL, Markus: Causal Inference in Sociological Research. In: *Annual Review of Sociology* (2010), Nr. 36, S. 21–47

[Gangl u. DiPete 2004] GANGL, Markus ; DIPETE, Thomas A.: Kausalanalyse durch Matchingverfahren. Version: 2004. http://opus.zbw-kiel.de/volltexte/2004/1830/pdf/dp401.pdf. In: DIEKMANN, Andreas (Hrsg.): *Methoden der Sozialforschung*. 2004

[Imbens u. Wooldridge 2009] IMBENS, Guido W. ; WOOLDRIDGE, Jeffrey M.: *Recent developments in the econometrics of program evaluation*. Estados Unidos : American Economic Association, 2009

[Legewie 2012] LEGEWIE, Joscha: Die Schätzung von kausalen Effekten: Überlegungen zu Methoden der Kausalanalyse anhand von Kontexteffekten in der Schule. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* (2012), Nr. 64, S. 123–153

[Manski 2003] MANSKI, Charles F.: *Partial identification of probability distributions*. New York : Springer, 2003 (Springer series in statistics). http://www.loc.gov/catdir/enhancements/fy0813/2003042476-b.html. – ISBN 0387004548

[Rubin 1986] RUBIN, Donald B.: Comment: Which Ifs Have Causal Answers: Comment: Which Ifs Have Causal Answers. In: *Journal of the American Statistical Association* (1986), S. 961–962

[Schuster 2009] SCHUSTER, Tibor: *Verteilungsbasierte kausale Inferenzmodelle zur Schätzung von Therapieeffekten in randomisierten, kontrollierten, klinischen Studien*. 2009

[Steyer 1992] STEYER, Rolf: *Theorie kausaler Regressionsmodelle: Vollst. zugl.: Trier, Univ., Habil.-Schr.* Stuttgart u.a : Fischer, 1992. – ISBN 3–437–50351–0

[Wolf 2010] WOLF, Christof (Hrsg.): *Handbuch der sozialwissenschaftlichen Datenanalyse*. 1. Aufl. Wiesbaden : VS Verl. für Sozialwissenschaften, 2010. – ISBN 978–3531163390