

Reinowski, Eva

**Working Paper**

## Mikroökonomische Evaluation und das Selektionsproblem – Ein anwendungsorientierter Überblick über nichtparametrische Lösungsverfahren –

IWH Discussion Papers, No. 200/2004

**Provided in Cooperation with:**

Halle Institute for Economic Research (IWH) – Member of the Leibniz Association

*Suggested Citation:* Reinowski, Eva (2004) : Mikroökonomische Evaluation und das Selektionsproblem – Ein anwendungsorientierter Überblick über nichtparametrische Lösungsverfahren –, IWH Discussion Papers, No. 200/2004, Leibniz-Institut für Wirtschaftsforschung Halle (IWH), Halle (Saale), <https://nbn-resolving.de/urn:nbn:de:gbv:3:2-5725>

This Version is available at:

<https://hdl.handle.net/10419/23738>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



**Mikroökonomische Evaluation  
und das Selektionsproblem**  
– Ein anwendungsorientierter Überblick  
über nichtparametrische Lösungsverfahren –

*Eva Reinowski*

Dezember 2004

Nr. 200

**Diskussionspapiere stehen in der alleinigen Verantwortung des jeweiligen Autors.  
Die darin vertretenen Auffassungen stellen keine Meinungsäußerung des IWH dar.**

**Diskussionspapiere**  
*Discussion Papers*

Autor: Eva Reinowski  
Abteilung Arbeitsmarkt  
Eva.Reinowski@iwh-halle.de  
Tel.: (0345) 7753-862

Diskussionspapiere stehen in der alleinigen Verantwortung des jeweiligen Autors. Die darin vertretenen Auffassungen stellen keine Meinungsäußerung des IWH dar.

Anregungen und kritische Bemerkungen sind jederzeit willkommen und erwünscht.

Herausgeber:

INSTITUT FÜR WIRTSCHAFTSFORSCHUNG HALLE (IWH)

Postanschrift: Postfach 11 03 61, 06017 Halle (Saale)

Hausanschrift: Kleine Märkerstraße 8, 06108 Halle (Saale)

Telefon: (03 45) 77 53-60

Telefax: (03 45) 77 53-8 20

Internet: <http://www.iwh-halle.de>

## **Zusammenfassung**

Im vorliegenden Diskussionspapier wird ein Überblick über nichtparametrische Verfahren zur Lösung des Selektionsproblems gegeben, wobei die bisher bestehende Lücke zwischen einführenden Texten und der weiterführenden, eher formalen Literatur geschlossen werden soll. Dazu werden die Vor- und Nachteile der Verfahren, die nötigen Annahmen sowie die Anforderungen an die Datenbasis erläutert. Aus der Darstellung wird ersichtlich, dass für die empirische Anwendung keine Methode zur Lösung des Selektionsproblems den anderen generell überlegen ist. Alle Ansätze beruhen auf bestimmten Grundannahmen, und jeder Schätzer ist verzerrt, wenn die zur Verfügung stehende Datenbasis nicht zu diesen Annahmen passt.



## 1 Einführung

Der Begriff Evaluation bezeichnet im allgemeinen den Vergleich einer eingetretenen Situation mit verschiedenen Alternativen. Ein solcher Vergleich ist immer dann sinnvoll, wenn aus verschiedenen Entscheidungsmöglichkeiten die beste ausgewählt werden soll bzw. im Nachhinein geprüft wird, ob die gewählte Alternative die beste (im Sinne eines bestimmten Kriteriums) gewesen ist. Insofern ist Evaluation auf verschiedenste Lebensbereiche bzw. ökonomische Entscheidungen anwendbar. Ein Beispiel, das gegenwärtig in der Anwendung solcher Verfahren dominiert, sind staatliche Fördermaßnahmen auf dem Arbeitsmarkt.

Für die Darstellungen im vorliegenden Diskussionspapier wird deshalb unterstellt, dass die untersuchte Handlungsalternative die Teilnahme an einer staatlicher Fördermaßnahme ist, deren Auswirkung auf die persönliche Situation anhand des Einkommenseffekts überprüft werden soll. Anstelle von Einkommenseffekten sind allerdings auch andere Zielgrößen, bspw. Arbeitslosigkeitsdauer, Wiederbeschäftigungschancen oder die Qualität einer künftigen Beschäftigung denkbar.

Die Schwierigkeit mikroökonomischer Evaluationsverfahren besteht darin, eine möglichst passende Vergleichsgröße zum beobachteten Einkommen zu finden. Das wird durch das sog. *Selektionsproblem* erschwert. Zu seiner Lösung wurden verschiedene parametrische und nichtparametrische Verfahren entwickelt. Nichtparametrische Lösungsverfahren sind besonders attraktiv, weil sie auch zu konsistenten Ergebnissen führen, wenn der Maßnahmeeffekt in der Bevölkerung heterogen ist.

Das vorliegende Diskussionspapier gibt einen Überblick über die nichtparametrischen Verfahren zur Lösung des Selektionsproblems. Dabei wird versucht, die bisher bestehende Lücke zwischen einführenden, eher intuitiven Texten und der weiterführenden, eher formalen Literatur zu schließen. So kann zum einen ein umfassender Überblick über die aktuell in der Literatur diskutierten Methoden gewonnen werden. Zum anderen kann der Neueinstieg in die Evaluationsproblematik erleichtert werden, indem für die einzelnen Verfahren folgende Fragen beantwortet werden: Welche Idee liegt dem jeweiligen Ansatz zugrunde? Welche Annahmen müssen erfüllt sein, damit das vorgestellte Verfahren konsistente Schätzergebnisse liefert? Welche Anforderungen werden an die Datenbasis gestellt? Welche Vorteile bietet die Methode? Welche Probleme können auftreten?

Die Beantwortung dieser Fragen soll außerdem zur Erleichterung der Auswahl des richtigen Verfahrens für die praktische Anwendung beitragen. Da die Gültigkeit der Grundannahmen der vorgestellten Verfahren i. d. R. nicht testbar ist, kann allerdings keine

allgemeingültige Handlungsempfehlung für die Verfahrensauswahl gegeben werden – außer dieser: jede Anwendung erfordert gründliche Vorüberlegungen und Sensibilität gegenüber der Kompatibilität von gewähltem Verfahren und der zur Verfügung stehenden Datenbasis.

Im folgenden Kapitel wird das Grundproblem der mikroökonomischen Evaluation erläutert, aus dessen Lösung das Problem der Selektionsverzerrung resultiert. Anschließend werden der Vorher-Nachher-Vergleich, der Kreuz-Vergleich und der Differenz-von-Differenzen-Schätzer vorgestellt. Gegenstand des vierten Kapitels sind verschiedene Matchingmethoden und mögliche Erweiterungen. Im letzten Kapitel werden die Hauptgedanken noch einmal zusammengefasst.

## 2 Das Grundproblem der mikroökonomischen Evaluation

Das Anliegen mikroökonomischer Evaluationen ist die Feststellung eventueller Veränderungen der persönlichen Situation der Betroffenen, bspw. durch Inanspruchnahme einer staatlichen Fördermaßnahme. Darüber kann ein Vergleich der persönlichen Situation, die aus dieser Entscheidung resultiert, mit derjenigen ohne Fördermaßnahme Aufschluss geben. Dabei wird unterstellt, dass die persönliche Situation ohne Teilnahme repräsentativ für die persönliche Situation ohne Maßnahme ist.<sup>1</sup>

Im Rahmen des *Model of Potential Outcomes*<sup>2</sup>, das als Grundlage jeder Evaluation verwendet wird, wird der individuellen Effekt einer Fördermaßnahme  $\alpha_{it}$  als Unterschied zwischen der potentiellen Situation einer Person  $i$  zum Zeitpunkt  $t$  nach Teilnahme an einer Fördermaßnahme  $Y_{it}^T$  und der potentiellen Situation ohne Maßnahmeteilnahme  $Y_{it}^C$  ermittelt:

$$(1) \quad \alpha_{it} = Y_{it}^T - Y_{it}^C .$$

---

<sup>1</sup> Diese Annahme impliziert, dass nur direkte Effekte auf die teilnehmenden Personen betrachtet werden und indirekte Effekte (z. B. Steuerzahlungen zur Finanzierung der Maßnahme oder Verdrängungseffekte auf dem regionalen Arbeitsmarkt) unberücksichtigt bleiben. Vgl. dazu Heckman et al. (1999), S. 1880.

<sup>2</sup> Dieses Modell findet sich in der Literatur unter verschiedenen Namen. Es wird auch als *Switching Regression Model* von Quandt, *Neyman-Fischer-Cox-Rubin-Model*, *Modell der potentiellen Einkommen* oder als *Roy-Modell der Einkommensverteilung* - siehe bspw. Heckman/Vytlacil (1996), S. 4730 für eine Verwendung dieser Begriffe – bzw. kurz als *Roy-Rubin-Modell* – siehe Hujer/Caliendo (2000), S. 9 - bezeichnet.

Die Verwendung des Modells basiert auf der Annahme, dass die Teilnahmeentscheidung von einem nutzenmaximierenden Individuum auf Grundlage seines individuellen erwarteten Nutzens aus der Maßnahme getroffen wird.<sup>3</sup>

Eine der beiden Situationen ist für ein und dieselbe Person allerdings nicht beobachtbar: für einen Teilnehmer lässt sich nicht feststellen, wie hoch bspw. sein Einkommen gewesen wäre, wenn er nicht an der Fördermaßnahme teilgenommen hätte; für eine nicht geförderte Person kann kein Einkommen nach Teilnahme an einer Maßnahme ermittelt werden.<sup>4</sup> Das beobachtete Einkommen  $Y_{it}$  ist entweder das Einkommen nach Teilnahme  $Y_{it}^T$  oder das Einkommen ohne Maßnahmeteilnahme  $Y_{it}^C$ :

$$(2) \quad Y_{it} = D_i \cdot Y_{it}^T + (1 - D_i) Y_{it}^C.$$

Die Dummy-Variable  $D_i$  gibt an, ob die betrachtete Person an einer Fördermaßnahme teilgenommen hat ( $D_i = 1$ ) oder nicht ( $D_i = 0$ ).

Das *Grundproblem der Evaluation* ist also ein Problem fehlender Informationen: der individuelle Effekt einer Fördermaßnahme ist nicht ermittelbar, da eine der beiden Situationen nicht beobachtbar ist und damit der Vergleichsmaßstab für die beobachtete Situation fehlt.<sup>5</sup> Anstelle des individuellen Effekts wird deshalb ein durchschnittlicher Maßnahmeeffekt bestimmt.<sup>6</sup>

Eine grundlegende Bedingung, die erfüllt sein muss, damit der Durchschnittseffekt unabhängig von Größe und Zusammensetzung der Teilnehmergruppe bestimmt werden kann, wird als *Stable Unit Treatment Value Assumption (SUTVA)* bezeichnet. Sie besagt, dass der Maßnahme-Effekt auf eine Person nicht beeinflusst werden darf von der Teilnahme anderer Personen an der gleichen Maßnahme. In der Praxis ist diese Bedin-

---

<sup>3</sup> Der Nutzen wird definiert als der Gegenwartswert der erwarteten Einnahmen nach Maßnahmeteilnahme, abzüglich der Teilnahmekosten (die sich aus direkten Kosten und entgangenen Einnahmen aufgrund der Teilnahme zusammensetzen). Individuelle Erwartungen werden unter Verwendung der dem Entscheidungsträger zur Verfügung stehenden Informationen gebildet und können daher – je nach Informationsquelle – zwischen den Teilnehmern variieren.

Für nähere Erläuterungen zur unterstellten Entscheidungsregel und den getroffenen Annahmen siehe Heckman et al. (1999), S. 1915f.

<sup>4</sup> Im folgenden werden die Begriffe „Einkommen“ und „Situation einer Person“ synonym verwendet. Das Einkommen ist als ein mögliches Beispiel für die Bewertung einer Situation zu verstehen. Die unbeobachtbare der beiden Situationen wird in der Literatur als *hypothetisch* bzw. *counterfactual (outcome)* bezeichnet. Vgl. bspw. Hujer/Caliendo (2000), S. 10.

<sup>5</sup> Das Fundamentalproblem wird häufig als *Identification Problem* bezeichnet. Vgl. Schmidt (1999), S. 3.

<sup>6</sup> Für eine Systematisierung möglicher Maßnahmeeffektschätzer siehe Hübler (2001), S. 104f.

gung nicht selbstverständlich erfüllt. So können bspw. große Teilnehmerzahlen an gleichartigen Förderprogrammen in einer Region die Beschäftigungsaussichten eines einzelnen Teilnehmers nach Beendigung einer Maßnahme negativ beeinflussen.

Für empirische Evaluationsstudien wird i. d. R. die Gültigkeit der *SUTVA* unterstellt.

Der prominenteste der durchschnittlichen Maßnahmeeffektschätzer ist der *durchschnittliche Effekt für die Teilnehmer*  $E(\alpha_i | D = 1)$ .<sup>7</sup> Er ist interpretierbar als Antwort auf die Frage „Wie groß ist der durchschnittliche Vorteil einer Person, der sich durch eine Maßnahme-Teilnahme ergibt, im Vergleich zu seiner Situation ohne Maßnahme?“ und wird ermittelt aus dem Unterschied der Situation der Teilnehmer nach der Maßnahme  $E(Y_i^T | D = 1)$  und der hypothetischen Situation, die sich eingestellt hätte, wenn diese Gruppe nicht teilgenommen hätte  $E(Y_i^C | D = 1)$ :

$$(3) \quad E(\alpha_i | D = 1) = E(Y_i^T | D = 1) - E(Y_i^C | D = 1).$$

Dieser in der englischsprachigen Literatur als *Treatment on the Treated* bezeichnete Schätzer erlaubt eine Schätzung des Maßnahmeeffekts ohne Einschränkung der Heterogenität der betrachteten Teilnehmer.

Da man für die Teilnehmergruppe keine Nichtteilnahme-Situation beobachten kann, muss für den letzten Term in Gleichung (3) ein geeigneter Ersatz gefunden werden: eine Gruppe nicht an der evaluierten Maßnahme Teilnehmender, für die gilt:

$$(A1) \quad E(Y_i^C | D = 1) - E(Y_i^C | D = 0) = 0.$$

Die Dummy-Variable  $D$  gibt dabei an, ob es sich um Maßnahmeteilnehmer ( $D = 1$ ) oder um Nichtteilnehmer ( $D = 0$ ) handelt.

Diese Annahme erfordert Gruppen, deren Mitglieder sich in allen für das Einkommen und die Maßnahme-Teilnahme relevanten Merkmalen<sup>8</sup> nicht unterscheiden. Dies gilt sowohl für beobachtbare Merkmale (wie Qualifikation, Alter, Geschlecht), als auch für unbeobachtbare (z. B. Motivation, Selbstbewusstsein).

In der Praxis ist diese Bedingung nicht selbstverständlich erfüllt, da die Teilnehmer an einer Fördermaßnahme i. d. R. eine ausgewählte Gruppe von Personen mit bestimmten

---

<sup>7</sup> Vgl. Schmidt (1999), S. 16.

<sup>8</sup> Relevante Merkmale sind solche, die sowohl die Zuordnung zu einer der beiden Gruppen als auch die potentielle Ausprägung der abhängigen Variable, also z. B. des Einkommens, beeinflussen. Vgl. dazu Eichler/Lechner (2001), S. 230.

Merkmale darstellen und deshalb nicht ohne weiteres mit den Nichtteilnehmern vergleichbar sind:<sup>9</sup>

$$E(Y_i^c | D = 1) - E(Y_i^c | D = 0) \neq 0.$$

Bei Nichtbeachtung der Unterschiede zwischen den Gruppen wird der Einfluss einer Maßnahme über- oder unterschätzt. Dieses Problem wird als *Selektionsverzerrung* bezeichnet.

Zur Beseitigung der Selektionsverzerrung wurden verschiedene Ansätze entwickelt, die sich hinsichtlich der Problemlösungsstrategie, der Anforderungen an die zur Verfügung stehenden Daten und der Möglichkeit, bestimmte Ursachen der Verzerrung zu berücksichtigen, unterscheiden. Im folgenden sollen Lösungsansätze erläutert werden, die keinen modelltheoretischen Rahmen erfordern.

### 3 Nichtparametrische Ansätze mit pauschaler Kontrollgruppenbildung

Die nichtparametrischen Ansätze können als alternative Entwicklung zu sozialen Experimenten innerhalb der statistischen Literatur angesehen werden.<sup>10</sup> Auch sie beruhen auf der Idee, das Problem der Selektionsverzerrung durch die Schaffung eines geeigneten Vergleichsmaßstabs zu beseitigen. Im Unterschied zu den experimentellen Ansätzen wird dabei allerdings keine neue Datenbasis geschaffen, sondern aus den vorliegenden Daten eine geeignete Kontrollgruppe konstruiert. Dabei kann entweder die gesamte zur Verfügung stehende Nichtteilnehmergruppe berücksichtigt oder ein Teil der Nichtteilnehmer für die Untersuchung ausgewählt werden. Entsprechend wird hier unterschieden in Lösungsansätze mit pauschaler Kontrollgruppenbildung und Lösungsansätze mit ausgewählten Nichtteilnehmern. Erstere lassen sich nach der Art der gewählten Kontrollgruppe in *Vorher-Nachher-Vergleich*, *Kreuz-Vergleich* und *Differenz-von-Differenzen-Schätzer* einteilen. Innerhalb der zweiten Gruppe werden verschiedene *Matchingverfahren* und mögliche Erweiterungen unterschieden.

<sup>9</sup> Das ergibt sich aus der Tatsache, dass für die einzelnen Förderprogramme bestimmte Zugangsvoraussetzungen bestehen, und nur diejenigen Personen, die diese Voraussetzungen erfüllen, an diesen Maßnahmen teilnehmen können.

<sup>10</sup> Eine ausführliche Darstellung experimenteller Ansätze findet sich in Rosenbaum (2002), Kapitel 2.

### 3.1 Der Vorher-Nachher-Vergleich

Der intuitiv am einfachsten zugängliche Maßnahmeeffekt-Schätzer vergleicht eine Personengruppe mit sich selbst.

#### 3.1.1 Idee und zentrale Annahme

Zur Ermittlung des Maßnahmeeffekts werden die Durchschnittseinkommen der Teilnehmergruppe vor und nach der Maßnahme miteinander verglichen.

Dazu wird unterstellt, dass sich das durchschnittliche Einkommen im Zeitablauf nicht verändern würde, wenn die betrachtete Personengruppe nicht an einer Fördermaßnahme teilnehmen würde:

$$(A2) \quad E(Y_t^c | D = 1) = E(Y_{t-a}^c | D = 1).$$

Dabei bezeichnen  $Y_{t-a}^c$  und  $Y_t^c$  die Einkommen vor Beginn bzw. nach Beendigung der Maßnahme.

Diese Kernannahme des Vorher-Nachher-Vergleichs lässt zwar Unterschiede zwischen Vor-Maßnahme-Einkommen und hypothetischem Einkommen für einzelne Personen zu, im Gruppendurchschnitt sind diese Abweichungen aber ausgeglichen.<sup>11</sup>

Die Annahme impliziert, dass das Einkommen vor Maßnahmebeginn nicht beeinflusst wird durch die Antizipation der Teilnahme an der bevorstehenden Maßnahme und dass keine gesamtwirtschaftlichen Veränderungen (wie Rezessionen oder Konjunkturschwankungen) und keine zeitbedingten Veränderungen (z. B. lebenszyklusabhängige Einkommenssteigerungen) das Durchschnittseinkommen der Teilnehmergruppe von einer Periode zur nächsten verändern.<sup>12</sup>

#### 3.1.2 Ermittlung des Maßnahmeeffekts

Die Gültigkeit der Kernannahme vorausgesetzt, kann der durchschnittliche Maßnahmeeffekt für die Teilnehmer  $\alpha_{BAE_t}$  zum Zeitpunkt  $t$  konsistent geschätzt werden:<sup>13</sup>

---

<sup>11</sup> Vgl. Schmidt (1999), S. 29.

<sup>12</sup> Vgl. Hujer et al. (2001), S. 7.

<sup>13</sup> Der *Vorher-Nachher-Vergleich* wird in der englischsprachigen Literatur als *Before After Estimator* bezeichnet. Daraus resultiert die hier gewählte Abkürzung  $\alpha_{BAE}$ .

$$(4) \quad \alpha_{BAE_t} = E\left[\left(Y_t^T \mid D = 1\right) - \left(Y_{t-a}^C \mid D = 1\right)\right].$$

Er ergibt sich als durchschnittliche Differenz der beobachteten Teilnehmereinkommen vor und nach Maßnahme.

Wenn die gesamte Teilnehmergruppe betrachtet wird, spricht man von einem *Naiven Vorher-Nachher-Vergleich*.<sup>14</sup> Die Berücksichtigung bestimmter erklärender Variablen wie z. B. Qualifikation, Alter oder Familienstand bei der Bildung der untersuchten Gruppe vergrößert die Chance, dass eventuelle Abweichungen zwischen Vor-Maßnahme-Einkommen und hypothetischem Einkommen im Gruppendurchschnitt tatsächlich beseitigt werden können.<sup>15</sup> Der Vorher-Nachher-Vergleich unter Berücksichtigung persönlicher Merkmale teilt die Gruppe der Teilnehmer in bestimmte Untergruppen von Personen, die jeweils die gleichen Merkmale haben.

Die Grundannahme muss dann für jede dieser Untergruppen erfüllt sein. Die berücksichtigten Merkmale werden mit  $X$  bezeichnet.

$$(A2') \quad E\left(Y_t^C \mid X, D = 1\right) = E\left(Y_{t-a}^C \mid X, D = 1\right)$$

Der durchschnittliche Einkommenseffekt einer Maßnahme  $\alpha_{BAE|X_t}$  zum Zeitpunkt  $t$  lässt sich dann in jeder Untergruppe als durchschnittliche Differenz der beobachteten Einkommen konsistent schätzen:<sup>16</sup>

$$(5) \quad \alpha_{BAE|X} = E\left[\left(Y_t^T \mid X, D = 1\right) - \left(Y_{t-a}^C \mid X, D = 1\right)\right].$$

Es sind verschiedene Darstellungsformen gebräuchlich. Alternativ zur Darstellung bedingter Erwartungswerte kann der Schätzer – in Anlehnung an Hübler (2001), S. 106 – auch wie folgt formuliert werden:

$$\alpha_{BAE_t} = \frac{1}{N_t} \sum_{i=1}^{N_t} \left(Y_{it}^T - Y_{i,t-a}^C\right).$$

Dabei bezeichnen  $Y_{it}^T$  und  $Y_{i,t-a}^C$  die individuellen Einkommen nach bzw. zu einem Zeitpunkt vor Beginn der Maßnahme.

<sup>14</sup> Diese Bezeichnung wird u. a. in Hujer/Caliendo/Radic (2001), S. 7 verwendet.

<sup>15</sup> Vgl. dazu Hujer et al. (2001), S. 7.

<sup>16</sup> In Analogie zum Naiven Vorher-Nachher-Vergleich lässt er sich alternativ darstellen als:

$$\alpha_{BAE|X} = \frac{1}{N_{t,X}} \sum_{i=1}^{N_{t,X}} \left(Y_{it}^T - Y_{i,t-a}^C\right).$$

Mit  $N_{t,X}$  wird dabei die Anzahl der Teilnehmer mit bestimmten Merkmalen  $X$  bezeichnet.

### 4.1.3 Kritik und Bemerkungen zur Anwendung

Ein großer Vorteil des Vorher-Nachher-Vergleichs besteht darin, dass keine Informationen über Nichtteilnehmer erforderlich sind. Es werden lediglich Längsschnitt-Daten der Teilnehmer für jeweils einen Zeitpunkt vor und nach der Maßnahme benötigt.<sup>17</sup> Die Anforderungen an die Datenbasis ist also vergleichsweise gering. Die Nutzung von ausschließlich Teilnehmer-Daten hat einen weiteren Vorteil: es können keine Verzerrungen durch heterogene beobachtbare und nicht beobachtbare Merkmale in den verglichenen Gruppen auftreten.

Allerdings ist die Grundannahme verletzt, wenn temporäre Veränderungen des Einkommens auftreten und diese nur die Teilnehmer betreffen. Ein Beispiel für eine solche Veränderung bildet der *Ashenfelter Dip*. Dieser Begriff beschreibt die Beobachtung, dass sich die Situation der zukünftigen Teilnehmer kurz vor Maßnahmebeginn verschlechtert.<sup>18</sup> Wenn der Referenzzeitpunkt in diesen Zeitraum fällt, wird der Maßnahmeeffekt überschätzt.

Eine weitere Verletzung der Grundannahme stellen zyklische Veränderungen der Arbeitsmarktsituation wie bspw. konjunkturelle Schwankungen dar. Eine arbeitsmarktbedingte Änderung der Situation eines Teilnehmers kann mit dem Vorher-Nachher-Vergleich nicht als solche identifiziert werden und führt – je nach Richtung der Veränderung - zur Über- oder Unterschätzung des Maßnahmeeffekts.

Simulationsstudien haben gezeigt, dass der Vorher-Nachher-Vergleich deutlich verzerrte Schätzergebnisse liefert, sobald gesamtwirtschaftliche Veränderungen, lebenszyklusbedingte Änderungen, Ashenfelter Dip oder eine Kombination aus mehreren „Störfaktoren“ auftritt.<sup>19</sup>

---

<sup>17</sup> In Heckman et al. (1999), S. 1892 wird darauf hingewiesen, dass Längsschnittdaten nicht zwingend erforderlich sind. Vorher-Nachher-Vergleiche sind auch mit wiederholten Querschnittsdaten möglich, wenn diese die gleiche Teilnehmergruppe repräsentieren und sich Abweichungen zwischen Vor-Maßnahme-Einkommen und hypothetischem Einkommen im Durchschnitt eliminieren lassen.

<sup>18</sup> Sehr anschaulich wird der Ashenfelter Dip und seine Auswirkungen auf Schätzergebnisse in Hujer et al. (2001), S. 7 f. erläutert. Ein häufig zitierter Erklärungsversuch dieses Phänomens von Ashenfelter (1978) besagt, dass infolge der Antizipation der Maßnahmeteilnahme die Suchanstrengungen der zukünftigen Teilnehmer nachlassen. Noch nicht geklärt ist allerdings, für welchen Zeitraum vor Maßnahmebeginn ein solches Verhalten tatsächlich unterstellt werden kann.

<sup>19</sup> Vgl. dazu bspw. Hujer et al. (2001), S. 14 ff.

Zum gleichen Ergebnis kommen auch Heckman/Smith (1999) beim Vergleich mit Ergebnissen eines sozialen Experiments. Vgl. S. 322 in Verb. mit Tab. 1.

Eine Möglichkeit der Kontrolle für arbeitsmarktbedingte Veränderungen bietet der Vergleich der Teilnehmergruppe mit einer geeigneten Gruppe von Nichtteilnehmern. Diese Idee liegt den im folgenden vorgestellten Methoden zugrunde.

### 3.2 Der Kreuz-Vergleich

Der einfachste Vertreter dieses Vergleichs zweier unterschiedlicher Personengruppen ist der *Kreuz-Vergleich*.

#### 3.2.1 Idee und zentrale Annahme

Hier werden die Einkommen von Teilnehmern und Nichtteilnehmern nach Beendigung der Maßnahme verglichen.

Dazu wird angenommen, dass das Durchschnittseinkommen in der Gruppe der Nichtteilnehmer mit dem hypothetischen Nichtteilnahme-Einkommen der Teilnehmer übereinstimmt.

Die unterschiedlichen Durchschnittseinkommen in beiden Gruppen resultieren also nur aus der Teilnahme der einen und der Nichtteilnahme der anderen Gruppe. Die Einkommen einzelner Personen in den Gruppen können sich unterscheiden, im Durchschnitt müssen sich die Unterschiede aber aufheben. Die durchschnittlichen Einkommen dürfen also weder durch beobachtbare noch durch unbeobachtbare Unterschiede in den Merkmalen der Personen beider Gruppen beeinflusst werden. Hier ergibt sich allerdings das Problem, dass Teilnehmer- und Kontrollgruppe aus sehr unterschiedlichen Personen bestehen können und damit nicht sichergestellt werden kann, dass das Einkommen in der Nichtteilnehmergruppe tatsächlich eine geeignete Vergleichsgrundlage bietet. Die Wahrscheinlichkeit der Einhaltung der Grundannahme kann daher durch die Berücksichtigung persönlicher Merkmale bei der Bildung der zu vergleichenden Gruppen erhöht werden.<sup>20</sup>

Dadurch entstehen mehrere Teilnehmer- und Nichtteilnehmergruppen mit Personen mit gleicher Ausprägung der Merkmale  $X$ .

Die Grundannahme lässt sich dann wie folgt formulieren:

$$(A3) \quad E(Y_t^C | X, D = 1) = E(Y_t^C | X, D = 0).$$

<sup>20</sup> Vgl. dazu bspw. Heckman et al. (1999), S. 1919.

Das hypothetische Einkommen in der Gruppe der Teilnehmer mit bestimmten Merkmalen  $E(Y_t^C | X, D = 1)$  stimmt mit dem Einkommen in der Gruppe der Nichtteilnehmer mit den gleichen Merkmalen  $E(Y_t^C | X, D = 0)$  überein.

### 3.2.2 Ermittlung des Maßnahmeeffekts

Der durchschnittliche Maßnahmeeffekt für die Teilnehmer  $\alpha_{CSE_t}$  zum Zeitpunkt  $t$  kann bei Gültigkeit dieser Annahme konsistent geschätzt werden mit Hilfe des Kreuz-Vergleichs:<sup>21</sup>

$$(6) \quad \alpha_{CSE_t|X} = E \left[ \left( Y_t^T | X, D = 1 \right) - \left( Y_t^C | X, D = 0 \right) \right].$$

Er ergibt sich aus der durchschnittlichen Differenz der Einkommen in Teilnehmer- und Nichtteilnehmergruppe nach Beendigung der Maßnahme.

### 3.2.3 Kritik und Bemerkungen zur Anwendung

Die Anwendung des Kreuz-Vergleichs erfordert Zugang zu Querschnittsdaten für die Maßnahmeteilnehmer und eine Gruppe nichtteilnehmender Personen für einen Zeitpunkt nach Beendigung der Maßnahme.

Im Gegensatz zum Vorher-Nachher-Vergleich ist er robust gegenüber gesamtwirtschaftlich und lebenszyklusbedingten Veränderungen, wenn diese die Teilnehmer und die Nichtteilnehmer gleichermaßen betreffen. Auch der Ashenfelter Dip stellt kein Problem dar, da nur Daten eines Zeitpunkts nach Maßnahmeende verwendet werden.

Wenn allerdings systematische Unterschiede zwischen den Durchschnittseinkommen beider Gruppen bestehen, bedingt durch beobachtbare oder unbeobachtbare Merkmale, liefert der Kreuz-Vergleich verzerrte Schätzergebnisse. Es ist beispielsweise denkbar, dass Motivationsunterschiede zwischen beiden Gruppen bestehen, die daraus resultieren, dass die aktiveren Personen sich mehr um eine Teilnahme an einem Förderprogramm bemühen und dadurch in der Teilnehmergruppe überrepräsentiert sind. Wenn

---

<sup>21</sup> In der englischsprachigen Literatur heißt dieser Schätzer *Cross Section Estimator* und wird deshalb hier mit  $\alpha_{CSE}$  bezeichnet.

Auch für diesen Schätzer gibt es eine alternative Darstellung :

$$\alpha_{CSE_t|X} = \frac{1}{N_{t,X}} \sum_{i=1}^{N_{t,X}} Y_{it}^T - \frac{1}{N_{c,X}} \sum_{j=1}^{N_{c,X}} Y_{jt}^C,$$

wobei mit  $N_{c,X}$  die Anzahl der Nichtteilnehmer mit bestimmten Merkmalen bezeichnet wird. Eine ähnliche Darstellung wird bspw. in Schmidt (1999), S. 30 oder Hübler (2001), S. 107 gewählt.

diese Personen schneller und/oder bessere Jobs bekommen, ist die Situation der Teilnehmergruppe auch ohne Maßnahme besser; der Maßnahmeeffekt würde überschätzt.

Unter bestimmten Annahmen ist ein anderes Verfahren, der *Differenz-von-Differenzen-Schätzer*, in der Lage, für systematische Unterschiede in nicht beobachteten Charakteristika zu kontrollieren.

### 3.3 Der Differenz-von-Differenzen-Schätzer

Diese Methode ist eine Kombination aus Vorher-Nachher-Vergleich und Kreuz-Vergleich.

#### 3.3.1 Idee und zentrale Annahme

Hier werden nicht Einkommen miteinander verglichen, sondern Einkommensveränderungen.

Es wird angenommen, dass die durchschnittliche Differenz zwischen Vor- und Nach-Maßnahme-Einkommen der Nichtteilnehmer der hypothetischen Einkommensdifferenz der Teilnehmer entspricht. Anders ausgedrückt:<sup>22</sup>

$$(A4) \quad E(Y_t^C - Y_{t-a}^C | X, D = 1) = E(Y_t^C - Y_{t-a}^C | X, D = 0).$$

Der Vorher-Nachher-Vergleich einer Nichtteilnehmergruppe mit bestimmten Merkmalen  $E(Y_t^C - Y_{t-a}^C | X, D = 0)$  wird als geeigneter Stellvertreter für den hypothetischen Vorher-Nachher-Vergleich der Teilnehmer mit den gleichen Merkmalen  $E(Y_t^C - Y_{t-a}^C | X, D = 1)$  angesehen.

Damit können die in den betrachteten Zeitpunkten  $t$  und  $t-a$  realisierten Einkommen zwischen Teilnehmern und Nichtteilnehmern differieren. Eine unverzerrte Schätzung des Maßnahmeeffekts ist allerdings nur möglich, wenn sich die Einkommensentwicklung in den verglichenen Personengruppen im Durchschnitt nur durch die Maßnahmeteilnahme der einen und die Nichtteilnahme der anderen Gruppe unterscheidet.

Unter welchen Bedingungen dies der Fall ist, lässt sich für einen „Durchschnittsteilnehmer“ und einen „Durchschnitts-Nichtteilnehmer“ mit Hilfe des Switching Reg-

---

<sup>22</sup> Wie für die oben vorgestellten Schätzer gilt auch hier, dass eine Berücksichtigung persönlicher Merkmale bei der Gruppenbildung die Wahrscheinlichkeit der Verzerrung der Schätzergebnisse verringert.

ression Model, das um eine unbeobachtbare individuelle Komponente  $\gamma_{it}$  erweitert wird, verdeutlichen:<sup>23</sup>

$$Y_{it} = \gamma_{it} + D_i Y_{it}^T + (1 - D_i) Y_{it}^C.$$

Unter Verwendung der Definition des individuellen Effekts einer Fördermaßnahme (1) lässt sich daraus die Differenz der Einkommen vor und nach der Maßnahme für den Teilnehmer  $i$  ableiten:

$$Y_{it} - Y_{i,t-a} = \gamma_{it} - \gamma_{i,t-a} + D_i \left[ (\alpha_{it} + Y_{it}^C) - Y_{i,t-a}^C \right].$$

Aus dem Vergleich mit der Einkommensdifferenz für den Nichtteilnehmer  $j$ :

$$Y_{jt} - Y_{j,t-a} = \gamma_{jt} - \gamma_{j,t-a} + (1 - D_j) \left[ Y_{jt}^C - Y_{j,t-a}^C \right]$$

lässt sich ablesen, wann die Differenz dieser Einkommensdifferenzen den Maßnahmeeffekt wiedergibt. Dazu müssen die Differenzen der Nichtteilnahmeeinkommen von Teilnehmer und Nichtteilnehmer gleich sein:  $(Y_{it}^C - Y_{i,t-a}^C) = (Y_{jt}^C - Y_{j,t-a}^C)$  und es muss möglich sein, den Einfluss der unbeobachtbaren Merkmale zu neutralisieren:  $(\gamma_{it} - \gamma_{i,t-a}) - (\gamma_{jt} - \gamma_{j,t-a}) = 0$ .

Dies ist unter folgenden Bedingungen der Fall:

- 1) wenn evtl. unbeobachtete Merkmale beider Personen und damit die unbeobachtbaren Heterogenitäten zeitinvariant sind:

$$\gamma_{it} = \gamma_{i,t-a} \quad \text{und} \quad \gamma_{jt} = \gamma_{j,t-a} \quad \text{sowie}$$

- 2) sich für die verglichenen Personengruppen die Nichtteilnahmeeinkommen im Durchschnitt gleich entwickeln:

$$\Delta Y_i^C = \Delta Y_j^C.$$

Das bedeutet auch, dass arbeitsmarkt- und lebenszyklusbedingte Einkommensveränderungen beide Gruppen gleichermaßen betreffen.

---

<sup>23</sup> Dabei ist zu berücksichtigen, dass die getroffenen Aussagen für den Gruppendurchschnitt gelten. Individuelle Abweichungen, die sich im Durchschnitt ausgleichen, sind zugelassen.

### 3.3.2 Ermittlung des Maßnahmeeffekts

Wenn die Grundannahme erfüllt ist, kann der durchschnittliche Einkommenseffekt einer Maßnahme für die Teilnehmer zum Zeitpunkt  $t$   $\alpha_{DiD_t|X}$  konsistent geschätzt werden:<sup>24</sup>

$$(7) \quad \alpha_{DiD_t|X} = E\left(Y_t^T - Y_{t-a}^C \mid X, D = 1\right) - E\left(Y_t^C - Y_{t-a}^C \mid X, D = 0\right).$$

Er wird ermittelt durch den Vergleich des Einkommensunterschieds der Teilnehmer mit bestimmten persönlichen Merkmalen  $E\left(Y_t^T - Y_{t-a}^C \mid X, D = 1\right)$  mit dem der Nichtteilnehmer mit den gleichen Merkmalen  $E\left(Y_t^C - Y_{t-a}^C \mid X, D = 0\right)$ . Der Einkommensunterschied bezeichnet jeweils die Differenz zwischen Vor- und Nach-Maßnahme-Einkommen.<sup>25</sup>

### 3.3.3 Kritik und Bemerkungen zur Anwendung

Die Differenz-von-Differenzen-Methode verbindet die Eigenschaften des Vorher-Nachher-Vergleichs mit denen des Kreuz-Vergleichs. Sie erfordert Zugang zu Längsschnittdaten für die Teilnehmer an der evaluierten Maßnahme und eine geeignete Gruppe von nicht teilnehmenden Personen.

Der Schätzer ist robust gegenüber gesamtwirtschaftlich und lebenszyklusbedingten Veränderungen, wenn diese die Teilnehmer und die Kontrollgruppenmitglieder gleichermaßen betreffen.<sup>26</sup> Er ist in der Lage, Selektionsverzerrung wegen unbeobachtbarer Heterogenitäten zu beseitigen, wenn sie durch zeitinvariante Charakteristika hervorgerufen wird.

Allerdings ist die Grundannahme verletzt, wenn unsymmetrische temporäre Veränderungen wie der Ashenfelter Dip auftreten. Die Ergebnisse sind also sehr sensitiv gegen-

<sup>24</sup> Der Schätzer wird wegen seiner englischen Bezeichnung *Difference in Difference Estimator* mit  $\alpha_{DiD_t}$  bezeichnet.

<sup>25</sup> Wie die oben beschriebenen Schätzer lässt auch er sich – in Anlehnung an Hübler (2001), S. 108 – alternativ darstellen als:

$$\alpha_{DiD_t|X} = \left[ \frac{1}{N_{t,X}} \sum_{i=1}^{N_{t,X}} Y_{it}^T - Y_{i,t-a}^C \right] - \left[ \frac{1}{N_{c,X}} \sum_{j=1}^{N_{c,X}} Y_{jt}^C - Y_{j,t-a}^C \right].$$

Mit  $Y_{j,t-a}^C$  wird das Einkommen der Nichtteilnehmer zu einem Zeitpunkt vor Beginn der Maßnahme bezeichnet.

<sup>26</sup> Wenn die Einkommensänderungen in beiden Gruppen unterschiedlich ausfallen, liefert die Differenz-von-Differenzen-Methode verzerrte Schätzergebnisse. Das wurde bei einem Vergleich der Ergebnisse dieser Methode mit denen eines sozialen Experiments nachgewiesen. Vgl. dazu Heckman/Smith (1999), S. 315.

über der Wahl der Zeitpunkte, zu denen das Vor- bzw. Nach- Maßnahme-Einkommen beobachtet wird.<sup>27</sup>

## 4 Nichtparametrische Ansätze mit ausgewählten Nichtteilnehmern

In den bisher vorgestellten Lösungsansätzen wird eine Kontrollgruppe für die beobachteten Teilnehmer aus allen Nichtteilnehmern, die bestimmte persönliche Merkmale haben, gebildet. Dabei ist uninteressant, ob wirklich für jeden Teilnehmer ein mit ihm übereinstimmender Nichtteilnehmer in der Kontrollgruppe zu finden ist. Eine andere Möglichkeit besteht darin, für jeden Teilnehmer einzeln einen passenden Nichtteilnehmer zu suchen und diese Nichtteilnehmer in der Kontrollgruppe zusammenzufassen. Das ist aufwendiger, hat aber den Vorteil, dass für jeden Teilnehmer ein Pendant in der Kontrollgruppe existiert und damit die Vergleichbarkeit der Gruppen verbessert wird.

Diese Idee der Zuordnung einzelner Nichtteilnehmer liegt dem *Matching-Ansatz* zugrunde.

### 4.1 Der Matching-Ansatz

Das Matching-Verfahren kann als eine Weiterentwicklung des Kreuz-Vergleichs betrachtet werden.

#### 4.1.1 Idee und Zentrale Annahme

Wie dort wird das hypothetische Einkommen der Teilnehmer durch das Einkommen geeigneter Nichtteilnehmer ersetzt. Im Unterschied zum Kreuz-Vergleich wird beim Matching allerdings für jeden einzelnen Teilnehmer eine eigene Unter-Kontrollgruppe ermittelt, für die gilt:

$$(A5) \quad E(Y_t^C | X, D = 1) = E(Y_t^C | X, D = 0).$$

Das durchschnittliche Einkommen der ausgewählten Nichtteilnehmer  $E(Y_t^C | X, D = 0)$  kann als Stellvertreter für das hypothetische Durchschnittseinkommen der Teilnehmer  $E(Y_t^C | X, D = 1)$  verwendet werden.

---

<sup>27</sup> Vgl. dazu bspw. Heckman/Smith (1999), S. 325.

Die Gültigkeit dieser Annahme setzt die Einhaltung folgender Bedingungen voraus. Die Grundbedingung, bezeichnet als *Conditional Independence Assumption (CIA)*, besagt, dass es möglich ist, unter Berücksichtigung der relevanten Merkmale Personen unabhängig vom potentiell erzielten Einkommen der Teilnehmer- oder Kontrollgruppe zuzuordnen.<sup>28</sup>

$$(A6) \quad Y_t^T, Y_t^C \perp D \mid X.$$

Anders ausgedrückt: Für alle Personen mit gleichen Merkmalen ist das potentielle Einkommen bei Teilnahme und Nichtteilnahme gleich.

Für die Ermittlung des durchschnittlichen Maßnahmeeffekts für die Teilnehmer muss diese Bedingung nur vom Nichtteilnahmeeinkommen erfüllt werden:<sup>29</sup>

$$(A6') \quad Y_t^C \perp D \mid X.$$

Dieser Annahme bedeutet, dass – gegeben  $X$  – die Nichtteilnahmeeinkommen in Teilnehmer- und Nichtteilnehmergruppe die gleiche Verteilung haben.<sup>30</sup> Das bedeutet, dass alle Merkmale, die den Selektionsprozess und das potentielle Einkommen beeinflussen, beobachtbar sind und berücksichtigt wurden.

Weiterhin müssen sowohl Teilnehmer als auch Nichtteilnehmer mit den beobachteten Merkmalen zu finden sein.

$$(A7) \quad 0 < \Pr(D = 1 \mid X) < 1.$$

Diese Annahme wird als *common support* bezeichnet.

Beide Bedingungen lassen sich unter dem Begriff *Strongly Ignorable Treatment Assignment (SITA)* zusammenfassen.<sup>31</sup>

---

<sup>28</sup>Für die Auswahl der relevanten Merkmale können theoretische Überlegungen mit den Ergebnissen früherer empirischer Studien verknüpft werden. Christensen (2001), S. 25 f. bspw. verbindet einen humankapitaltheoretischen Ansatz mit früheren Forschungsergebnissen, um Variablen zu identifizieren, die die Teilnahmeentscheidung an Weiterbildungsmaßnahmen beeinflussen.

<sup>29</sup> Vgl. Heckman et al. (1997), S. 611.

<sup>30</sup> Vgl. Heckman et al. (1998), S. 1023.

<sup>31</sup> Vgl. Rosenbaum/Rubin (1983), S. 43.

#### 4.1.2 Ermittlung des Maßnahmeeffekts

Der durchschnittliche Maßnahmeeffekt für die Teilnehmer  $\alpha_{M_t}$  zum Zeitpunkt  $t$  lässt sich bei Gültigkeit der oben genannten Annahmen – analog zum Kreuz-Vergleich – konsistent aus der durchschnittlichen Einkommensdifferenz zwischen Teilnehmern und Nichtteilnehmern ermitteln:

$$(8) \quad \alpha_{M_t} = E \left[ \left( Y_t^T \mid X, D=1 \right) - \left( Y_t^C \mid X, D=0 \right) \right].$$

Der Unterschied zum *Kreuz-Vergleich* wird aus der alternativen Darstellung des Schätzers ersichtlich. Anstelle des Vergleichs zweier Durchschnittseinkommen wird hier der Maßnahmeeffekt als gewichteter Durchschnitt aller individuellen Einkommensdifferenzen ermittelt:

$$(8') \quad \alpha_{M_t} = \sum_{i=1}^{N_T} w(i) (Y_{it}^T - \bar{Y}_{it}^C).$$

Die Anzahl der berücksichtigten Einkommensdifferenzen wird dabei mit  $N_T$  bezeichnet, die Gewichtung jeder einzelnen Differenz mit  $w(i)$ .

$\bar{Y}_{it}^C$  gibt das Vergleichseinkommen für Teilnehmer  $i$  zum Zeitpunkt  $t$  an, das aus dem gewichteten Durchschnitt aller Nichtteilnehmereinkommen in seiner Unter-Kontrollgruppe gebildet wird:

$$(9) \quad \bar{Y}_{it}^C = \sum_{j=1}^{N_{C_i}} W(i, j) Y_{jt}^C.$$

Mit  $Y_{jt}^C$  wird das Einkommen eines Nichtteilnehmers  $j$  zum Zeitpunkt  $t$  bezeichnet;  $W(i, j)$  gibt die individuelle Gewichtung dieses Einkommens für Teilnehmer  $i$  an. Die einzelnen Nichtteilnehmereinkommen können unterschiedlich gewichtet werden. Für die Gewichtung gilt:

$$(10) \quad 0 \leq W(i, j) \leq 1 \text{ und}$$

$$\sum_{j=1}^{N_{C_i}} W(i, j) = 1.$$

Die Anzahl der Nichtteilnehmer in der Unter-Kontrollgruppe  $C_i$  wird durch  $N_{C_i}$  angegeben.

Wie groß die Unter-Kontrollgruppe jedes Teilnehmers ist und wie die einzelnen Einkommen gewichtet werden, hängt von der Wahl des Zuordnungsprozesses ab. Wichtig ist dabei, dass diejenigen Nichtteilnehmer, die dem Teilnehmer besonders ähnlich sind, auch am stärksten berücksichtigt werden.

#### 4.1.3 Matching-Zuordnungsprozesse

Die Unter-Kontrollgruppe  $C_i$  eines Teilnehmers wird in Abhängigkeit seiner individuellen Merkmale  $X_i$  geschaffen und setzt sich aus denjenigen Nichtteilnehmern zusammen, deren Charakteristika  $X_j$  denen des Teilnehmers ähnlich sind:

$$(11) \quad C_i = \{j \mid X_j \in C(X_i)\} \quad \forall j \in \{D = 0\}.$$

Mit  $C(X_i)$  wird die Menge aller Nichtteilnehmer mit ähnlichen Merkmalen<sup>32</sup> bezeichnet.

Die Zuordnungsprozesse lassen sich danach unterscheiden, ob genau ein Nichtteilnehmer ausgewählt wird, oder ob mehrere Nichtteilnehmer in der Unter-Kontrollgruppe vertreten sind.

Zuordnungsprozesse, in denen genau ein Nichtteilnehmer die Unter-Kontrollgruppe bildet, werden als *1:1-Matching* oder *Nearest-Neighbor-Matching* bezeichnet und sollen im folgenden Abschnitt erläutert werden. Andere Verfahren, in denen mehrere Nichtteilnehmer zur Konstruktion des Vergleichseinkommens verwendet werden, sind Gegenstand der darauf folgenden Abschnitte.

##### 4.1.3.1 Nearest-Neighbor-Matching

Für das *1:1-Matching* ergibt sich folgende Definition der Ähnlichkeit zwischen dem Teilnehmer und den Nichtteilnehmern in der Unter-Kontrollgruppe:

$$(11.1) \quad C(X_i) = \left\{ j \mid \underset{j \in \{D=0\}}{\text{Min}} \| X_i - X_j \| \right\}.$$

Dabei bezeichnet  $\| X_i - X_j \|$  die Distanz zwischen einem Teilnehmer und einem Nichtteilnehmer, die aufgrund der beobachteten Merkmale ermittelt wird. Es wird derjenige Nichtteilnehmer als ähnlichster bezeichnet, dessen Merkmale den geringsten Ab-

---

<sup>32</sup> In der Literatur findet sich für Ähnlichkeit auch der Begriff *Neighborhood*. Vgl. bspw. Heckman et al. (1999), S. 1953.

stand zu den Teilnehmermerkmalen aufweisen. Für die Gewichtung des Nichtteilnehmereinkommens gilt dann:

$$(10.1) \quad W(i, j) = \begin{cases} 1 & \text{wenn } j \in C(X_i) \\ 0 & \text{sonst.} \end{cases}$$

Das Teilnehmereinkommen wird nur mit dem Einkommen des ihm ähnlichsten Nichtteilnehmers verglichen.

Um zu verhindern, dass ein Teilnehmer mit einem ihm unähnlichen Nichtteilnehmer verglichen wird, wenn es keine ähnlichen Nichtteilnehmer gibt, kann ein zulässiger Maximalabstand zwischen den Merkmalen festgelegt werden. Diese Form wird als *Caliper Matching* bezeichnet:<sup>33</sup>

$$(11.2) \quad C(X_i) = \begin{cases} \left\{ j \mid \underset{j \in \{D=0\}}{\text{Min}} \| X_i - X_j \| \right\} & \text{wenn } \| X_i - X_j \| < \varepsilon \\ \emptyset & \text{sonst.} \end{cases}$$

Es wird dann unter denjenigen Nichtteilnehmern der ähnlichste ausgewählt, deren Merkmale nur innerhalb der vorgegebenen Toleranz  $\varepsilon$  von denen des Teilnehmers abweichen. Existiert kein solcher Nichtteilnehmer, kann für den betrachteten Teilnehmer kein Vergleichseinkommen ermittelt werden; er wird bei der Ermittlung des Maßnahmeeffekts nicht berücksichtigt.

Für die Feststellung des Abstands zwischen Teilnehmer- und Nichtteilnehmermerkmalen lassen sich die gängigen statistischen Abstandsmaße verwenden.<sup>34</sup> Ein weit verbreitetes Maß ist die *Mahalanobis Metrik*  $MM_{ij}$

$$(12) \quad MM_{ij} = \left[ (X_i) - (X_j) \right]' \Sigma_{C_{ij}}^{-1} \left[ (X_i) - (X_j) \right],$$

bei der neben dem quadrierten Abstand jedes einzelnen Merkmals die Varianz-Kovarianz-Matrix der Merkmale  $\Sigma_{C_{ij}}$  berücksichtigt wird. Das hat den Vorteil, dass evtl.

---

<sup>33</sup> Vgl. Cochrane/Rubin (1973), S. 420 f.

<sup>34</sup> Eine Einführung in die vor allem in der Clusteranalyse eingesetzten Abstandsmaße bietet jedes statistische Lehrbuch, bspw. Backhaus (2000), S. 265 ff.

vorhandenen Korrelationen zwischen den Merkmalen Rechnung getragen wird.<sup>35</sup> Der ähnlichste Nichtteilnehmer würde mit Hilfe dieses Maßes wie folgt definiert:

$$(11.3) \quad C(X_i) = \left\{ j \mid \underset{j \in \{D=0\}}{\text{Min}} \| MM_{ij} \| \right\}.$$

Die Anwendung dieses Distanzmaßes ist allerdings nur möglich, wenn alle berücksichtigten Merkmale metrisch skaliert sind. Sollen unterschiedlich skalierte Merkmale in die Distanzmessung einbezogen werden, ist die Bildung eines aggregierten Distanzmaßes erforderlich. Allerdings ist die Suche nach einem geeigneten Gewichtungsfaktor, der die einzelnen Merkmale entsprechend ihrer Bedeutung berücksichtigt, sehr aufwendig. Hierfür existieren keine verbindlichen Richtlinien.<sup>36</sup>

Eine Alternative zur Bildung eines aggregierten Distanzmaßes ist die Abstandsmessung mit Hilfe eines *Balancing Scores*. Unter diesem Begriff werden alle Funktionen  $b(X)$  der relevanten Merkmale zusammengefasst, für die gilt:

$$(A8) \quad X \perp D \mid b(X).$$

Gegeben die Funktion  $b(X)$ , muss die bedingte Verteilung der Merkmale in Teilnehmer- und der Kontrollgruppe gleich sein.<sup>37</sup> Also ermöglicht auch die Verwendung eines *Balancing Scores* in großen Stichproben die Einordnung von Personen in Teilnehmer- oder Kontrollgruppe unabhängig vom potentiell erzielten Einkommen.<sup>38</sup>

Der größte – und in empirischen Studien am häufigsten verwendete – *Balancing Score* ist der *Propensity Score*  $P(X)$ .<sup>39</sup> Er kann als die auf Grundlage der relevanten Merkmale geschätzte Wahrscheinlichkeit der Teilnahme an einer Fördermaßnahme aufgefasst werden:  $P(X) = \Pr(D = 1 \mid X)$ .<sup>40</sup>

<sup>35</sup> Vgl. Kaufmann/Pape (1996), S. 450.

<sup>36</sup> In der gängigen statistischen Literatur wird eine Gewichtung der für jedes Skalenniveau einzeln ermittelten Distanzmaße entsprechend des Anteils der Variablen an der Gesamtvariablenanzahl vorgeschlagen. Vgl. bspw. Opitz (1980), S. 50 f. Um eine Über- oder Untergewichtung einer der Variablengruppen zu verhindern, müssen die Variablen vorher normiert werden.

<sup>37</sup> Die Definition des *Balancing Score* in dieser Form findet sich in Rosenbaum/Rubin (1983), S. 42.

<sup>38</sup> In Rosenbaum/Rubin (1983), S. 45 wird bewiesen, dass eine Zuordnung auf Grundlage eines *Balancing Scores* die Grundannahme des Matchings (SITA) asymptotisch erfüllt, wenn der Score aus Merkmalen gebildet wird, für die diese Annahme erfüllt ist. Es gilt also:

$$Y_i^T, Y_i^C \perp D \mid b(X) \text{ und } 0 < \Pr(D = 1 \mid b(X)) < 1.$$

<sup>39</sup> Anwendungen finden sich bspw. bei Rosenbaum/Rubin (1985), Hujer et al. (1999), Hujer/Wellner (2000), Larsson (2000) oder Brodaty et al. (2001).

<sup>40</sup> Durch die Schätzung wird der Rahmen der nichtparametrischen Verfahren eigentlich verlassen, denn mit der Wahl eines Modells wird eine implizite Verteilungsannahme getroffen.

Für die Schätzung der individuellen Teilnahmewahrscheinlichkeit wird i. d. R. ein *Probit-Modell* verwendet. Der Modellteil der latenten Variable  $I_i$  wird als *Indexfunktion* bezeichnet:

$$I_i = \beta X_i + \varepsilon_i \quad \text{mit } \varepsilon_i \sim N(0, \sigma^2),$$

der beobachtbare Modellteil, die Teilnahmeentscheidung, wird beschrieben durch:

$$D_i = \begin{cases} 1 & \text{wenn } I_i > 0 \\ 0 & \text{sonst.} \end{cases}$$

Der individuelle *Propensity Score* lässt sich in diesem Modellrahmen unter Verwendung der Verteilungsfunktion der Standardnormalverteilung  $\Phi$  schätzen:<sup>41</sup>

$$(13) \quad P(X_i) = \Phi(\hat{\beta} X_i).$$

Der Grund für die häufige Verwendung dieses eindimensionalen Abstandsmaßes liegt in der erheblichen Vereinfachung des Zuordnungsprozesses, da nur die Ähnlichkeit dieser einen Maßzahl anstelle aller relevanten Merkmale überprüft werden muss – womit das *Dimensionsproblem*<sup>42</sup> umgangen wird. Allerdings ist zu berücksichtigen, dass die Einhaltung der Grundannahme nur asymptotisch gilt. Gerade in kleinen Stichproben kann der mit der Verwendung jedes eindimensionalen Ähnlichkeitsmaßes verbundene Informationsverlust Auswirkungen auf die Qualität des Matchingergebnisses haben. So ist nicht auszuschließen, dass einige Merkmale eine andere Bedeutung für die Teilnahmeentscheidung als für die zukünftige Situation einer Person haben, sodass Personen mit identischem Propensity Score unterschiedliche Aussichten auf dem Arbeitsmarkt haben. Diesem Problem kann begegnet werden, indem – zusätzlich oder alternativ zum Propen-

---

<sup>41</sup> Für eine detaillierte Erläuterung des *Probit-Modells* und weiterer Annahmen vgl. Greene (2003), Kapitel 21.

<sup>42</sup> Mit diesem Begriff wird die exponentielle Steigerung der Anzahl der zu überprüfenden möglichen Übereinstimmungen zwischen Teilnehmern und potentiellen Partnern, die mit jedem zusätzlich berücksichtigten Merkmal verbunden ist, bezeichnet.

In Anlehnung an Hujer et al. (2001), S. 10 lässt sich dieses Problem mit Hilfe eines Beispiels verdeutlichen: Die Ähnlichkeit von Personen soll anhand von  $n=2$  Merkmalen überprüft werden, für die jeweils 2 Ausprägungen zu beobachten sind (z. B. Beschäftigungsstatus: beschäftigt/nicht beschäftigt; abgeschlossene Berufsausbildung: ja/nein). Es ergeben sich  $2^2=2^n$  zu überprüfende Matching-Möglichkeiten. Die Berücksichtigung eines zusätzlichen Merkmals (Alter: bis 45 Jahre/älter) erhöht diese Anzahl auf  $2^3=2^{n+1}$ .

sity Score – wichtige Determinanten der persönlichen Situation bei der Feststellung der Ähnlichkeit von Personen berücksichtigt werden.<sup>43</sup>

Die oben beschriebenen Methoden zur Ähnlichkeitsfeststellung lassen sich miteinander kombinieren. Eine sehr häufig verwendete Kombination, die von Rosenbaum/Rubin (1985) vorgeschlagen und von Lechner (1998) weiterentwickelt wurde, verbindet den Propensity Score als Vorauswahlkriterium mit einem Mahalanobis-Metrik-Matching zur Ermittlung der Ähnlichkeit zwischen Teilnehmern und den vorgewählten Nichtteilnehmern.<sup>44</sup>

Im ersten Schritt wird für jeden Teilnehmer eine Gruppe von Nichtteilnehmern ausgewählt. Als Kriterium dient die aus der Propensity Score-Schätzung abgeleitete Teilnahmeneigung<sup>45</sup> des Teilnehmers:

$$(14) \quad \hat{I}_i = \hat{\beta} X_i.$$

Es werden nur diejenigen Nichtteilnehmer als Mitglieder der vorausgewählten Vergleichsgruppe  $C_{i,VOR}$  für Teilnehmer  $i$  zugelassen, deren Teilnahmeneigung  $\hat{I}_j$  nur innerhalb der Grenzen eines Intervalls von der des Teilnehmers abweicht:

$$(15) \quad C_{i,VOR} = \left\{ j \mid \hat{I}_j \in \left[ \hat{I}_i \pm \varepsilon \sqrt{\text{Var}(\hat{I}_i)} \right] \right\} \quad \forall j \in \{D = 0\}.$$

Die Größe des Intervalls ist abhängig von der Genauigkeit der Schätzung: je geringer die Streuung, desto kleiner das Intervall und umso ähnlicher damit die Mitglieder dieser Vergleichsgruppe. Die Konstante  $\varepsilon$  wird dabei so gewählt, dass ein Konfidenzintervall um die Teilnahmeneigung des Teilnehmers entsteht.

Für den Teilnehmer und jeden Nichtteilnehmer in der Vergleichsgruppe  $j \in C_{i,VOR}$  wird dann die Ähnlichkeit anhand aller relevanten Merkmale mit Hilfe der Mahalanobis Metrik festgestellt:

$$(16) \quad MM_{ij} = \left[ \left( \hat{I}_i, X_i \right) - \left( \hat{I}_j, X_j \right) \right]' \Sigma_{C_j}^{-1} \left[ \left( \hat{I}_i, X_i \right) - \left( \hat{I}_j, X_j \right) \right].$$

<sup>43</sup> Dieses Problem wird in Fröhlich (2004), S. 216f. ausführlich erläutert. Der aus der zusätzlichen Berücksichtigung persönlicher Merkmale resultierende *Balancing Score* wird hier als *Augmented Propensity Score* bezeichnet.

<sup>44</sup> Beim Vergleich der Ergebnisse ihres zweistufigen Verfahrens mit denen des *Nearest Neighbor Propensity Score Matching* und des *Nearest Neighbor Mahalanobis Metrik Matching* stellten Rosenbaum/Rubin (1985), S. 37 ff. eine größere Ähnlichkeit zwischen Teilnehmern und den zugeordneten Partnern als in beiden Vergleichsverfahren fest.

<sup>45</sup> Die Bezeichnung der latenten Variable der Probit-Schätzung als Teilnahmeneigung ist übernommen aus Lechner (1998), wo das hier erläuterte zweistufige Verfahren zum ersten Mal beschrieben ist.

Als ähnlichster Nichtteilnehmer wird derjenige mit der geringsten Mahalanobis Distanz zum Teilnehmer definiert:

$$(11.4) \quad C(X_i) = \left\{ j \mid \underset{j \in C_{i,VOR}}{\text{Min}} \|MM_{ij}\| \right\}.$$

Die Zuordnungsprozesse, mit denen den Teilnehmern ihre Partner zugewiesen werden, lassen sich einteilen in Zuordnungsprozesse nach dem Prinzip *Ziehen mit Zurücklegen* und solche nach dem Prinzip *Ziehen ohne Zurücklegen*.

Bei der *Zuordnung mit Zurücklegen* wird jedem Teilnehmer der ähnlichste Nichtteilnehmer zugewiesen, unabhängig davon, für wie viele Teilnehmer ein Nichtteilnehmer als Partner verwendet wird. Dabei besteht die Gefahr, dass nur wenige Nichtteilnehmer zur Bildung der Kontrollgruppe benutzt werden, auch wenn noch andere – sehr ähnliche – Nichtteilnehmer zur Verfügung stehen.<sup>46</sup>

Die Anwendung dieser Zuordnung empfiehlt sich für Matchingprozesse, bei denen Teilnehmer- und Nichtteilnehmergruppe (annähernd) gleich groß sind, da sichergestellt ist, dass jeder Teilnehmer zur Feststellung des Maßnahmeeffekts berücksichtigt wird, wenn er ein Pendant in der Nichtteilnehmergruppe hat.<sup>47</sup>

Wenn die Nichtteilnehmergruppe deutlich größer als die Teilnehmergruppe ist, werden *Zuordnungsprozesse ohne Zurücklegen* angewendet. Bei dieser auch als *Pair Matching* bezeichneten Art der Gruppenbildung wird jedem Teilnehmer genau ein Nichtteilnehmer zugeordnet, wobei ein Nichtteilnehmer nicht als Partner für mehrere Teilnehmer eingesetzt werden kann.

Ein häufig eingesetztes Verfahren legt nach dem Zufallsprinzip eine Reihenfolge fest, nach der jedem Teilnehmer ein passender Nichtteilnehmer zugeordnet wird. Diejenigen Teilnehmer, für die kein Partner gefunden wird, werden aus der Gruppe der untersuchten Teilnehmer entfernt.<sup>48</sup> Die implizite Antwort auf das Entscheidungsproblem, welchem Teilnehmer ein Nichtteilnehmer zugeordnet werden soll, der als Partner für meh-

---

<sup>46</sup> Lechner (2000), S. 9 weist darauf hin, dass damit eine erhebliche Steigerung der Varianz der Schätzung verbunden sein kann.

<sup>47</sup> Gleich große Teilnehmer- und Nichtteilnehmergruppen treten vor allem bei der Evaluation mehrerer Maßnahmen auf, wo die Teilnehmer der einen Maßnahme gleichzeitig Nichtteilnehmer für eine andere Maßnahme sind. Vgl. bspw. Fröhlich (2002), Fröhlich et al. (2000), Lechner (2001).

<sup>48</sup> In Augurzky (2000) findet sich für dieses Verfahren der Begriff *Greedy Pair Matching*. Für die Anwendung dieses Verfahrens vgl. bspw. Lechner (1998), Christensen (2001) oder Gerfin/Lechner (2002).

rere Teilnehmer in Frage kommt, lautet hier: Dem zufällig zuerst gezogenen. Dabei wird in Kauf genommen, dass – je nachdem, ob ein Toleranzbereich für die Abweichung der Merkmale festgelegt wurde oder nicht – bei der Zuordnung evtl. Teilnehmer übrig bleiben, weil kein passender Nichtteilnehmer mehr gefunden wird, obwohl es ähnliche Nichtteilnehmer gibt, oder diesen Teilnehmern sehr unähnliche Nichtteilnehmer zugeordnet werden.<sup>49</sup>

Eine bessere Antwort wäre also: Demjenigen Teilnehmer, der keinen weiteren möglichen Partner hat. Damit könnte verhindert werden, dass mehr Teilnehmer als nötig aus der Analyse ausgeschlossen werden.

Auch wenn der Verlust von Teilnehmern für die Analyse keine Rolle spielen würde, gäbe es eine bessere Entscheidung: Die Zuordnung des dem Nichtteilnehmer ähnlichen Teilnehmers. Damit würde die Ähnlichkeit zwischen Teilnehmer- und Kontrollgruppe insgesamt erhöht.<sup>50</sup>

Die gleichzeitige Berücksichtigung beider Anliegen ist mit Hilfe *optimaler Matching-Prozesse* möglich. Dazu können Methoden aus der *linearen Optimierung* bzw. der *Graphentheorie* eingesetzt werden.<sup>51</sup> Solche Verfahren liefern eine vollständige Zuordnung von Nichtteilnehmern zu Teilnehmern (wenn diese möglich ist), bei der die Gesamtsumme der Abstände zwischen Teilnehmern und Nichtteilnehmern minimal ist.<sup>52</sup>

Die Gefahr aller hier beschriebenen *Zuordnungsprozesse ohne Zurücklegen* besteht darin, dass der festgestellte Maßnahmeeffekt nicht repräsentativ für die gesamte Gruppe der Teilnehmer ist. Dies ist der Fall, wenn Teilnehmer mit ganz bestimmten Merkmalen systematisch aus der Gruppe der untersuchten Personen entfernt werden, weil kein passender Nichtteilnehmer gefunden wird.<sup>53</sup> Diesem Problem kann auf verschiedene Weise begegnet werden.

---

<sup>49</sup> Auf diese Problematik wird in Rosenbaum (1989), S. 1026 f. anhand von Beispielen eingegangen.

<sup>50</sup> Die Anwendung eines iterativen Matching-Algorithmus, bei dem die nach dem Standardverfahren gebildete Anfangszuordnung verändert wird, wenn sich dadurch die Ähnlichkeit zwischen beiden Gruppen insgesamt verbessert, findet sich in Reinowski/Schultz/Wiemers (2003).

<sup>51</sup> Die Zusammenhänge zwischen einem Matching-Problem und der Graphentheorie werden in Rosenbaum (1989), S. 1027 ff. ausführlich erläutert.

<sup>52</sup> Die Anwendung solcher Zuordnungsprozesse auf Matching-Probleme ist allerdings relativ selten. Ein Anwendungsbeispiel des *Ungarischen Algorithmus* zur Evaluation von Weiterbildungsmaßnahmen findet sich bspw. in Reinowski/Schultz/Wiemers (2004).

<sup>53</sup> Vgl. dazu Augurzky (2000), S. 3.

#### 4.1.3.2 Matching mit mehreren Nichtteilnehmern und Full Matching

Eine Möglichkeit der Zuordnung besteht in der Aufteilung aller Nichtteilnehmer unter den Teilnehmern. Bei diesem als *Matching mit mehreren Nichtteilnehmern* bezeichneten Verfahren wird eine Mindestanzahl von Partnern, die jedem Teilnehmer zugeordnet werden müssen, festgelegt (i. d. R. einer). Auch die Festlegung von Höchstgrenzen ist möglich, um ausgewogene Gruppengrößen zu erreichen. Die Unter-Kontrollgruppen der Teilnehmer setzen sich dann aus unterschiedlich vielen Nichtteilnehmern zusammen. Für die Größe der Unter-Kontrollgruppe jedes Teilnehmers gilt:  $a \leq N_{C_i} \leq b$ . Dabei bezeichnet  $N_{C_i}$  die Anzahl der Nichtteilnehmer in der Kontrollgruppe,  $a$  gibt die minimale Anzahl der Partner eines Teilnehmers,  $b$  die maximal mögliche Anzahl an.<sup>54</sup>

Für die Feststellung der Distanz zwischen Teilnehmern und Nichtteilnehmern können die oben beschriebenen statistischen Abstandsmaße eingesetzt werden. Die Zuordnung der Nichtteilnehmer erfolgt mit Hilfe von Algorithmen aus der linearen Optimierung. Dabei werden die Personen so zugeordnet, dass die Gesamtsumme der Distanzen über alle Teilnehmer und Nichtteilnehmer minimiert wird – unter Berücksichtigung der Grenzen für die Gruppengrößen.

Für die Gewichtung der einzelnen Nichtteilnehmereinkommen zur Konstruktion der Vergleichsgröße sind verschiedene Möglichkeiten denkbar. Beispielsweise lässt sich das Vergleichseinkommen als arithmetisches Mittel der Nichtteilnehmereinkommen in einer Untergruppe darstellen. Jedes einzelne Einkommen wird dann wie folgt gewichtet:

$$(10.2) \quad W(i, j) = \frac{1}{N_{C_i}}.$$

Eine andere Möglichkeit ist die Gewichtung der Nichtteilnehmereinkommen in Abhängigkeit von der Distanz der Nichtteilnehmer vom Teilnehmer. Denkbar wäre bspw. die Verwendung bekannter Kernfunktionen wie *Gaußscher* oder *Epanechnikov-Kernel*.

Bei *Full Matching Verfahren* können mehrere Nichtteilnehmer einem Teilnehmer und mehrere Teilnehmer einem Nichtteilnehmer zugeordnet werden. Auch dabei werden alle Nichtteilnehmer verwendet. Die Untergruppen werden so gebildet, dass die durchschnittliche Distanz zwischen Teilnehmern und Nichtteilnehmern innerhalb der gebilde-

---

<sup>54</sup> Diese Möglichkeit der Zuordnung wird ausführlich in Rosenbaum (1989), S. 1029 f. beschrieben.

ten Untergruppen minimiert wird.<sup>55</sup> Auch hierfür lassen sich Verfahren der linearen Optimierung anwenden. Für die Ermittlung der Distanz zwischen Teilnehmern und Nichtteilnehmern und die Gewichtung der einzelnen Nichtteilnehmereinkommen gelten die oben getroffenen Aussagen.

In den beiden zuletzt beschriebenen Zuordnungsprozessen wird die komplette Nichtteilnehmergruppe für die Schätzung des Maßnahmeeffekts verwendet. Deshalb ist – zusätzlich zum durchschnittlichen Maßnahmeeffekt für die Teilnehmer – auch die Ermittlung des *durchschnittlichen Maßnahmeeffekts für eine zufällig ausgewählte Person* möglich.<sup>56</sup>

Wie für das Nearest Neighbour Matching ist auch für diese Zuordnungsverfahren die Festlegung von maximalen Distanzen, die zwischen einzelnen Personen bestehen dürfen, vorstellbar (Caliper Matching). Damit kann verhindert werden, dass völlig unähnliche Nichtteilnehmer bei der Konstruktion der Vergleichseinkommen berücksichtigt werden.<sup>57</sup> Allerdings kann dann nicht ausgeschlossen werden, dass für manche Teilnehmer kein passender Nichtteilnehmer gefunden wird, und diese damit nicht mehr zur Schätzung des Maßnahmeeffekts zur Verfügung stehen.

#### 4.1.4 *Local Polynomial Regression (Kernel Matching und Local Linear Matching)*

Ein alternativer Ansatz verwendet die Einkommen mehrerer Nichtteilnehmer zur Schätzung eines Vergleichseinkommens mit Hilfe einer gewichteten Kleinst-Quadrate-Schätzung (weighted least squares). Das Einkommen eines Nichtteilnehmers kann dabei für die Schätzung mehrerer Vergleichsgrößen genutzt werden.<sup>58</sup>

Für die Anwendung der *Local Polynomial Regression* wird angenommen, dass das Vergleichseinkommen für eine Umgebung von  $X_i$  durch ein Polynom  $p$ -ter Ordnung approximiert werden kann. Dann ist die Schätzung der Vergleichsgröße als gewichteter

---

<sup>55</sup> In Rosenbaum (1991), S. 600 f. wird gezeigt, dass jedes Full Matching die optimale Untergruppenbildung hinsichtlich der durchschnittlichen Distanz zwischen Teilnehmern und Nichtteilnehmern in einer Untergruppe repräsentiert.

<sup>56</sup> Augurzky (2000), S. 13

<sup>57</sup> Wenn dabei systematisch Nichtteilnehmer aus der weiteren Analyse ausgeschlossen werden, ist der Maßnahmeeffekt nicht mehr repräsentativ für eine zufällig ausgewählte Person, sondern für die Teilnehmergruppe.

<sup>58</sup> Dieses Vorgehen wird bspw. in Fröhlich (2000), S. 4 beschrieben.

Durchschnitt der Einkommen  $Y_j$  aller Nichtteilnehmer mit Merkmalen  $X_j$ , die den Merkmalen des betrachteten Teilnehmers  $X_i$  ähnlich sind, möglich. Die Gewichtung der einzelnen Nichtteilnehmereinkommen erfolgt so, dass die  $Y_j$  mit zunehmendem Abstand der  $X_j$  von  $X_i$  geringeres Gewicht bekommen.

Die Schätzung der Koeffizienten der Weighted Least Squares Regression entspricht der Lösung des folgenden Minimierungsproblems:

$$(17) \quad \min_{\beta_0, \dots, \beta_p} \sum_{j \in \{D=0\}} \left( Y_j - \sum_{l=0}^p \beta_l (X_i - X_j)^l \right)^2 K \left( \frac{X_i - X_j}{h_n} \right)$$

für  $l = 0, \dots, p$ . Dabei bezeichnet  $p$  die polynomiale Ordnung des Minimierungsproblems,  $K(\cdot)$  eine nichtnegative Gewichtungs- bzw. Kernfunktion und  $h_n$  die Bandbreite, innerhalb derer die Merkmale der berücksichtigten Nichtteilnehmer liegen müssen.<sup>59</sup>

Die Lösung lässt sich wie folgt formulieren:

$$(9') \quad \hat{Y}_i^C = \sum_{j \in \{D=0\}} W^n \left( \frac{X_i - X_j}{h_n} \right) Y_j$$

mit Gewichtung  $W^n \left( \frac{X_i - X_j}{h_n} \right) = e^T S_n^{-1} \left\{ 1, (X_i - X_j), \dots, (X_i - X_j)^p \right\}^T K \left( \frac{X_i - X_j}{h_n} \right)$ ,

wobei der Index  $T$  jeweils für die transponierte Matrix steht.

Die  $n \times (p+1)$ -Matrix  $S_n = \begin{pmatrix} S_{n,0}(X_j) & S_{n,1}(X_j) & \dots & S_{n,p}(X_j) \\ S_{n,1}(X_j) & S_{n,2}(X_j) & \dots & S_{n,p+1}(X_j) \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,p}(X_j) & S_{n,p+1}(X_j) & \dots & S_{n,2p}(X_j) \end{pmatrix}$  besteht aus den

Elementen  $S_{n,l}(X_j) = \sum_{j \in \{D=0\}} K \left( \frac{X_i - X_j}{h_n} \right) (X_i - X_j)^l$  (für  $l = 0, \dots, 2p$ ). Die Anzahl der

Elemente des Einheitsvektors  $e^T = (1, 0, 0, \dots)$  werden durch  $p$  vorgegeben.

<sup>59</sup> Es ist ersichtlich, dass auch dieser Ansatz nicht im wahrsten Sinne des Wortes nichtparametrisch ist, weil mit der Wahl der Gewichtungsfunktion implizit eine Annahme über die Verteilung der Einkommen getroffen wird.

Durch die Wahl der polynomialen Ordnung für die Approximation des Vergleichseinkommens wird die angewendete Methode festgelegt. So lässt sich bspw. die *Kernel-Regression* als Lösung der Local Polynomial Regression für  $p = 0$  darstellen:

$$(9'.1) \quad \hat{Y}_{i,0}^C = \sum_{j \in \{D=0\}} \frac{K\left(\frac{X_i - X_j}{h_n}\right)}{\sum_{j \in \{D=0\}} K\left(\frac{X_i - X_j}{h_n}\right)} Y_j$$

Die Wahl von  $p = 1$  impliziert die Anwendung der *Local Linear Regression*. Die Lösung dieses Verfahrens lässt sich im oben beschriebenen Rahmen wie folgt darstellen:

$$(9'.2) \quad \hat{Y}_{i,1}^C = \left( \frac{S_{n,2}(X_j)T_{n,0}(X_j) - S_{n,1}(X_j)T_{n,1}(X_j)}{S_{n,0}(X_j)S_{n,2}(X_j) - S_{n,1}^2(X_j)} \right),$$

wobei gilt:  $S_{n,l}(X_j) = \sum_{j \in \{D=0\}} K\left(\frac{X_i - X_j}{h_n}\right) (X_i - X_j)^l$  für  $l = 0, 1, 2$  und

$$T_{n,l}(X_j) = \sum_{j \in \{D=0\}} K\left(\frac{X_i - X_j}{h_n}\right) (X_i - X_j)^l Y_j \text{ für } l = 0, 1.$$

Polynome höherer Ordnung sind zur Lösung des Selektionsproblems nicht gebräuchlich.

Der Unterschied zwischen beiden Methoden wird deutlich, wenn man sich die gewichtete Regression zur Schätzung der Vergleichsgrößen vorstellt: die Regression für Kernel Matching entspricht der Schätzung einer Konstanten, während die Local Linear Regression eine Konstante und einen linearen Term enthält.<sup>60</sup>

#### 4.1.4.1 Kernel Matching

Beim Kernel Matching werden alle Nichtteilnehmereinkommen zur Konstruktion der Vergleichsgröße eingesetzt. Die Unter-Kontrollgruppe eines Teilnehmers entspricht also der gesamten Kontrollgruppe:

$$(11.5) \quad C(X_i) = j \in \{D = 0\}.$$

<sup>60</sup> Vgl. Heckman/Smith (1999), S. 342.

Die Gewichtung  $W(i, j)$  jedes einzelnen Kontrollgruppenmitglieds  $j$  für den betrachteten Teilnehmer  $i$  ist abhängig von der Ähnlichkeit seiner Merkmale zu denen des Teilnehmers. Sie ist nicht auf Null oder eins beschränkt:

$$(10.3) \quad W(i, j) = \frac{K_{ij}}{\sum_{k \in \{D=0\}} K_{ik}}.$$

Dabei wird mit  $h_n$  die gewählte Bandbreite und mit  $K_{ik} = K\left(\frac{X_k - X_i}{h_n}\right)$  die Kernfunktion bezeichnet, die zur Gewichtung der einzelnen Nichtteilnehmereinkommen für Teilnehmer  $i$  eingesetzt werden soll. Die Gewichtung der einzelnen Nichtteilnehmereinkommen richtet sich nach ihrem Abstand vom Teilnehmer. Als Kernfunktion kann jede symmetrische, positive Funktion, deren Integral eins ist, verwendet werden.<sup>61</sup> Die gebräuchlichsten Funktionen sind die Normalverteilung (Gaußscher Kernel) und der Epanechnikov-Kernel.

Durch die Wahl der Bandbreite wird festgelegt, wie viele Nichtteilnehmereinkommen für die Schätzung der Vergleichsgröße für einen Teilnehmer genutzt werden. Dabei besteht ein Zielkonflikt zwischen Unverzerrtheit und Varianz der Schätzung: Mit zunehmender Größe der Bandbreite verringert sich die Varianz des Schätzers (weil mehr Nichtteilnehmereinkommen zur Schätzung verwendet werden), aber die Gefahr der Verzerrung nimmt zu (weil auch solche Nichtteilnehmereinkommen verwendet werden, die dem Teilnehmer weniger entsprechen). Die optimale Bandbreite (im Sinne einer Minimierung des *mean square error*) ist abhängig von verschiedenen Faktoren und kann mit Hilfe verschiedener Methoden näherungsweise bestimmt werden. Die am häufigsten eingesetzten Verfahren sind *Silvermans Faustregel*<sup>62</sup> und das *Cross-Validation-Verfahren*, bei dem der durchschnittliche Schätzfehler minimiert wird.

#### 4.1.4.2 Local Linear Regression

Eine Verringerung des Bias kann erreicht werden, wenn ein Polynom höherer Ordnung zur Approximation des Vergleichseinkommens verwendet wird. Die Local Linear Regression hat neben ihren asymptotischen Eigenschaften (Minimax-Effizienz unter allen

---

<sup>61</sup> Diese Anforderungen werden von jeder symmetrischen Verteilungsfunktion erfüllt.

<sup>62</sup> Die Anwendung dieser Regel erfordert die Verwendung der *Gaußschen Kernfunktion*. Eine Anwendung findet sich bspw. in Bergemann et al. (2001), S. 7 ff.

linearen Schätzern)<sup>63</sup> weitere Vorzüge wie Robustheit gegenüber der Verteilung der Regressoren und damit Anwendbarkeit auf verschiedene Datendesigns, Algorithmen mit schneller Konvergenz der Lösung sowie einfache Interpretation der Ergebnisse.<sup>64</sup> Ein weiterer Vorteil besteht darin, dass (bspw. im Gegensatz zum Nadaraya-Watson-Kernel) keine Modifikationen für die Ränder der Verteilung nötig sind.

Auch für die Anwendung der Local Linear Regression wird die Vergleichsgröße aus den Einkommen aller Nichtteilnehmer konstruiert:

$$(11.6) \quad C(X_i) = j \in \{D = 0\}.$$

Für die Gewichtung  $W(i, j)$  jedes einzelnen Kontrollgruppenmitglieds  $j$  für den betrachteten Teilnehmer  $i$  gilt:

$$(10.4) \quad W(i, j) = \frac{K_{ij} \sum_{k \in \{D=0\}} K_{ik} (X_i - X_k)^2 - [K_{ij} (X_i - X_j)] \left[ \sum_{k \in \{D=0\}} K_{ik} (X_i - X_k) \right]}{\sum_{j \in \{D=0\}} K_{ij} \sum_{k \in \{D=0\}} K_{ik} (X_i - X_k)^2 - \left[ \sum_{k \in \{D=0\}} K_{ik} (X_i - X_k) \right]^2}.$$

Dabei bezeichnet  $K_{ik} = K\left(\frac{X_k - X_i}{h_n}\right)$  die Kernfunktion und  $h_n$  die Bandbreite. Für die

Wahl der Kernfunktion gelten die oben getroffenen Aussagen. Die Festlegung der Bandbreite erfolgt i. d. R. mit Hilfe des Cross-Validation-Verfahrens.

Bei der Anwendung der Local Linear Regression auf kleine Stichproben kann in Datenbereichen, in denen wenige Beobachtungen zu finden sind, das Problem auftreten, dass die Varianz der Schätzung sehr groß ist. Um die Varianz in diesen Bereichen zu verringern, wurden verschiedene Modifikationen entwickelt. Beim sog. *Trimming* werden nur Daten von Personen zur Schätzung verwendet, deren geschätzte Teilnahmewahrscheinlichkeit hinreichend häufig auftritt.<sup>65</sup> Dazu wird ein Bereich der geschätzten Dichtefunktion der Teilnahmewahrscheinlichkeit definiert, innerhalb dessen die geschätzte Teilnahmewahrscheinlichkeit der Personen liegen muss. Daten von Personen, deren geschätzte Wahrscheinlichkeitsdichte geringer ist, werden bei der Schätzung der Koeffizienten ausgeschlossen. Unter *Ridging* ist die „Stabilisierung“ der Schätzung durch

<sup>63</sup> Als Minimax-Effizient werden Schätzer bezeichnet, bei denen die maximal mögliche durchschnittliche Abweichung des Schätzergebnisses vom wahren Wert geringer ist als bei allen anderen Schätzern. Diese Eigenschaft wurde in Fan et al. (1997) bewiesen.

<sup>64</sup> Vgl. Hengartner et al. (2002), S.791 oder Heckman et al. (1998), S1041.

<sup>65</sup> In Heckman et al. (1998), S. 1078 f. wird diese Idee im Rahmen eines Partial Linear Regression Model vorgestellt und erläutert.

einen Zusatzterm zu verstehen. Der eingefügte *Ridge Parameter* vergrößert den Nenner der Gleichung (9'.2), wenn  $S_{n,2}(X_j)$  sehr klein ist, d. h., wenn sich die Beobachtungen im Intervall auf eine Stelle konzentrieren. Dadurch wird die Varianz der Schätzung verringert.<sup>66</sup>

#### 4.1.5 Kritik und Bemerkungen zur Anwendung

Matching ist das am häufigsten angewendete Verfahren zur Lösung des Selektionsproblems.

Diese Popularität lässt sich mit verschiedenen Vorzügen des Verfahrens erklären. Zum einen lässt sich die Grundidee der individuellen Suche nach Vergleichspersonen leicht vermitteln. Dadurch ist es besonders geeignet als empirische Grundlage der Politikberatung. Zum anderen ist Matching robust gegenüber gesamtwirtschaftlich und lebenszyklusbedingten Veränderungen, wenn diese die Teilnehmer und die Nichtteilnehmer gleichermaßen betreffen.<sup>67</sup> Wenn der Zeitpunkt der Beobachtung der Merkmale, aufgrund derer Teilnehmer und Nichtteilnehmer gematcht werden, lange genug vor Beginn der evaluierten Maßnahme liegt, kann auch der Einfluss eines *Ashenfelter Dip* ausgeschlossen werden.

Allerdings sind die Anforderungen an die Datenbasis sehr hoch: Es sind umfangreiche Informationen über Teilnehmer und Nichtteilnehmer für mindestens zwei Zeitpunkte nötig: vor Maßnahmebeginn und nach Ende der Maßnahme.<sup>68</sup> Für die Anwendung einer Nearest Neighbour Zuordnung muss zudem die Anzahl der Nichtteilnehmer, für die diese Informationen vorliegen, die Zahl der teilnehmenden Personen deutlich übersteigen, damit für jeden Teilnehmer ein geeigneter Nichtteilnehmer gefunden werden kann.<sup>69</sup> Aus diesem Grund ist in jüngeren Veröffentlichungen zunehmend die Anwendung von Verfahren aus der Gruppe der Local Polynomial Regression zu beobachten.

Der entscheidende Nachteil aller hier vorgestellten Matching-Verfahren besteht darin, dass mit ihrer Hilfe nur Heterogenitäten zwischen Teilnehmern und Nichtteilnehmern aufgrund beobachtbarer Merkmale beseitigt werden können. Es besteht also die Gefahr, dass das Selektionsproblem nicht vollständig gelöst wird, da die Heterogenitäten aufgrund unbeobachtbarer Eigenschaften unberücksichtigt bleiben.

---

<sup>66</sup> In Seifert/Gasser (2000), S. 340 ff. wird die Konstruktion des Ridge Parameters ausführlich erläutert.

<sup>67</sup> Vgl. dazu Schmidt (1999), S. 27.

<sup>68</sup> Vgl. dazu Hujer et al. (2001), S. 20.

<sup>69</sup> Vgl. dazu Christensen (2001), S. 28.

Um dieses Problem zu lösen, werden in der Literatur verschiedene Erweiterungen des Matching-Ansatzes diskutiert. Im Folgenden soll ein kurzer Überblick über die Verknüpfungsmöglichkeiten von Matching mit anderen Verfahren gegeben werden.

## 4.2 Erweiterungen des Matching-Ansatzes

### 4.2.1 Der bedingte Differenz-von-Differenzen-Schätzer

Eine Erweiterungsmöglichkeit besteht in der Kombination von Matching mit der Differenz-von-Differenzen-Methode. Dadurch können noch verbleibende Heterogenitäten aufgrund unbeobachtbarer Faktoren beseitigt werden, falls sie sich im Zeitablauf nicht ändern.<sup>70</sup>

Für die Schätzung des durchschnittlichen Maßnahmeeffekts wird das Teilnehmereinkommen  $Y_{it}^C$  durch die Einkommensdifferenz des Teilnehmers zwischen zwei Zeitpunkten  $t$ ,  $t-a$  ersetzt:  $\Delta Y_i^T = (Y_{it}^T - Y_{i,t-a}^C)$ . Die Vergleichsgröße bildet der gewichtete Durchschnitt der Einkommensdifferenzen zwischen zwei Zeitpunkten in der Unter-Kontrollgruppe  $\Delta \bar{Y}_i^C = \sum_{j=1}^{N_{C_i}} W(i, j) (Y_{it}^C - Y_{i,t-a}^C)$ . Der Maßnahmeeffekt wird dann ermittelt als gewichteter Durchschnitt der Differenzen aus den individuellen Einkommensdifferenzen der Teilnehmer:

$$(8'.1) \quad \alpha_{cDiD_t} = \sum_{i=1}^{N_T} w(i) [\Delta Y_i^T - \Delta \bar{Y}_i^C].$$

Die Anzahl der Teilnehmer wird mit  $N_T$  bezeichnet, die Gewichtung jedes Teilnehmers mit  $w(i)$ .

### 4.2.2 Regression-adjusted Matching

Ein alternativer Ansatz erweitert das oben vorgestellte Matchingverfahren um die Nutzung von Informationen über die funktionale Form der Einkommensgleichung und die

---

<sup>70</sup> Erläuterungen zu dieser Erweiterungsmöglichkeit finden sich in Heckman et al. (1997), S. 613 f. und 631 sowie in Heckman/Smith (1999), S. 342. Angewendet wird diese Kombination bspw. in Bergemann et al. (2001, 2004).

Einbeziehung von sog. *Ausschlussbedingungen*, d. h. der Zerlegung von  $X$  in verschiedene Komponenten.<sup>71</sup>

Den Ausgangspunkt bildet ein herkömmliches Modell für das Einkommen im Nichtteilnahmefall:  $Y_0 = X\beta_0 + U_0$ , in dem beobachtbare und unbeobachtbare Faktoren additiv separierbar sind. Zur Schätzung der Parameter wird eine partiell lineare Regression auf die Daten der Kontrollgruppe angewendet:  $E(Y_0 | X, D = 0) = X\beta_0 + E(U_0 | X, D = 0)$ . Dabei muss gelten:<sup>72</sup>

$$(A9) \quad U_0 \perp D | X.$$

Es ist in vielen Fällen möglich und sinnvoll,  $X$  in zwei Komponenten zu zerlegen:  $X = (T, Z)$ , in Variablen, die das Einkommen beeinflussen  $T$  und welche, die die Teilnahmewahrscheinlichkeit beeinflussen  $Z$ . Dabei können einige Variablen sowohl in  $T$  als auch in  $Z$  enthalten sein. Dann gilt:  $\Pr(D = 1 | X) = \Pr(D = 1 | Z)$  und

$$(A9') \quad U_0 \perp T | D, Z.$$

Für die Schätzung des Maßnahmeeffekts wird dann von  $Y_0$  und  $Y_1$  der Schätzer  $X\hat{\beta}_0$  abgezogen:  $\tilde{Y}_{it}^T = Y_{it}^T - X_{it}\hat{\beta}_0$  bzw.  $\tilde{Y}_{jt}^C = Y_{jt}^C - X_{jt}\hat{\beta}_0$ . Der durchschnittliche Maßnahmeeffekt wird unter den o. g. Bedingungen konsistent ermittelt als:<sup>73</sup>

$$(8'2) \quad \alpha_{RM_t} = \sum_{i=1}^{N_T} w(i) (\tilde{Y}_{it}^T - \tilde{Y}_{it}^C)$$

wobei die Vergleichsgröße bezeichnet wird mit:<sup>74</sup>

$$(9'') \quad \tilde{Y}_{it}^C = \sum_{j=1}^{N_C} W(i, j) \tilde{Y}_{jt}^C.$$

Beide Erweiterungen lassen sich miteinander verbinden zum sog. *Regression-adjusted conditional Difference-in-Difference-Verfahren*.<sup>75</sup>

<sup>71</sup> Die Grundidee dieser Methode wird in Heckman/Ichimura/Tod (1997), S. 630 f. erläutert.

<sup>72</sup> Diese Bedingung bedeutet nicht, dass  $E(U_0 | X) = 0$ , sondern die gleiche Verteilung der unbeobachtbaren Faktoren in Teilnehmer- und Nichtteilnehmergruppe. Vgl. Heckman/Ichimura/Todd (1998), S. 268.

<sup>73</sup> Vgl. Heckman et al. (1998), S. 269.

<sup>74</sup> Diese Idee lässt sich mit Local Linear Matching auf Grundlage des Propensity Scores verbinden. Anwendungen dieser Kombination finden sich bspw. in Heckman et al. (1997) und Heckman et al. (1998).

## 5 Schlussbetrachtung

Im vorliegenden Diskussionspapier werden verschiedene nichtparametrische Methoden zur Lösung des Problems der Selektionsverzerrung vorgestellt. Zunächst wird dazu das Grundproblem der Evaluierung und das daraus erwachsende Problem der Selektionsverzerrung erläutert. Im Anschluss daran wird die Lösung mit Hilfe verschiedener Verfahren der pauschalen Kontrollgruppenbildung sowie mit ausgewählten Nichtteilnehmern dargestellt. Es werden Vor- und Nachteile der Verfahren, die nötigen Annahmen und Anforderungen an die Datenbasis erläutert.

Aus der Darstellung ist ersichtlich, dass in der empirischen Anwendung keine Methode zur Lösung des Selektionsproblems den anderen generell überlegen ist.<sup>76</sup> Alle Ansätze beruhen auf bestimmten Grundannahmen über Teilnahmeentscheidung und Einkommen. Da diese Annahmen i. d. R. nicht testbar sind, sind gründliche Vorüberlegungen und eine genaue Analyse der Aussagefähigkeit der Daten für die Auswahl eines adäquaten Schätzverfahrens nötig. Denn jeder Schätzer liefert inkonsistente Ergebnisse, wenn die Datenbasis nicht zu den getroffenen Annahmen passt.<sup>77</sup>

In der Praxis begrenzt die zur Verfügung stehende Datenbasis die Auswahlfreiheit zwischen verschiedenen Modellen und auch die Güte der Kontrollgruppe, die konstruiert werden kann.<sup>78</sup> Die Antwort auf die Frage, welches Verfahren in der empirischen Anwendung eingesetzt werden sollte, ist also abhängig von Art, Umfang und Qualität der zur Verfügung stehenden Daten.

„The best solution to the evaluation problem lies in improving the quality of the data on which evaluations are conducted and not in the development of formal econometric methods to circumvent inadequate data.”<sup>79</sup>

---

<sup>75</sup> Vgl. Heckman et al. (1997), S. 631.

<sup>76</sup> Diese Aussage wird in Heckman et al. (1999), S. 201 ff. mit Hilfe einer Simulationsstudie untermauert.

<sup>77</sup> Vgl. Heckman et al. (1999), S. 2007.

<sup>78</sup> Vgl. dazu LaLonde (1986), S. 612.

<sup>79</sup> Siehe Heckman et al. (1999), S. 1869.

## Literatur

- ASHENFELTER, O. (1978): "Estimating the Effect of Training Programs on Earnings", *Review of Economics and Statistics*, 60, 47-57.
- AUGURZKY, B. (2000): "Evaluation Strategies in Labor Economics – An Application to Post-secondary Education", *Dissertation an der Ruprecht-Karls-Universität Heidelberg*
- BACKHAUS, K. (2000): "Multivariate Analysemethoden", 9. überarb und erw. Aufl., Berlin: Springer.
- BERGEMANN, A., B. FITZENBERGER und S. SPECKESSER (2001): "Evaluating the Employment Effects of Public Sector Sponsored Training in East Germany: Conditional Difference-in-Differences and Ashenfelter Dip", *Discussion Paper*, Mimeo, Mannheim, [http://www.vwl.uni-mannheim.de/ls\\_oe/forschung\\_index.htm](http://www.vwl.uni-mannheim.de/ls_oe/forschung_index.htm).
- BERGEMANN, A., B. FITZENBERGER und S. SPECKESSER (2004): "Evaluating the Dynamic Employment Effects of Training Programs in East-Germany Using Conditional Difference-in-Difference", *ZEW-Discussion Paper No. 04-41*, Zentrum für Europäische Wirtschaftsforschung GmbH, Mannheim.
- BRODATY, T., B. CREPON, und D. FOUGÈRE (2001): "Using Matching Estimators to Evaluate Alternative Youth Employment Programs: Evidence from France, 1986-1988", in: Lechner, M. und F. Pfeiffer, Hrsg.: *Econometric Evaluation of Labour Market Policy*, Heidelberg: Physica/Springer, 85-124.
- CHRISTENSEN, B. (2001): "Berufliche Weiterbildung und Arbeitsplatzrisiko: Ein Matching-Ansatz", *Kieler Arbeitspapier Nr. 1033*, Institut für Weltwirtschaft, Kiel.
- COCHRAN, W. und D. RUBIN (1973): "Controlling Bias in Observational Studies: A Review", *Sakhyā*, Ser. A, 35, 417-446.
- EICHLER, M. und M. LECHNER (2001): "Public Sector Sponsored Continuous Vocational Training in East Germany: Institutional Arrangements, Participants and Results of Empirical Evaluations", in: Riphahn, R., D. Snower und K. Zimmermann, Hrsg.: *Employment Policy in Transition: The Lessons from German Integration for the Labour Market*, Berlin: Springer, 208-253.
- FAN, J., T. GASSER, I. GIJBELS, M. BROCKMANN, und J. ENGEL, (1997): "Local Polynomial Regression: Optimal Kernels and Asymptotic Minimax Efficiency" *Annals of the Institute of Mathematical Statistics*, 49, 79-99.
- FRÖHLICH, M. (2004): "Programme Evaluation with Multiple Treatments", *Journal of Economic Surveys*, 18, 181-224.
- FRÖHLICH, M. (2000): "Treatment Evaluation: Matching versus Local Polynomial Regression", *Discussion Paper No. 2000-17*, Department of Economics, University of St. Gallen, St. Gallen.

- FRÖHLICH, M., A. HESHMATI und LECHNER, M. (2000): "Mikroökonomische Evaluierung berufsbezogener Rehabilitation in Schweden", *Swiss Journal of Economics and Statistics*, 136, 433-461.
- GERFIN, M. und M. LECHNER (2002): "A Microeconomic Evaluation of the Active Labour Market Policy in Switzerland", *The Economic Journal*, 482, 854-893.
- GREENE, W. (2003): "Econometric Analysis", 5. Aufl., New Jersey: Prentice Hall.
- HECKMAN, J., H. ICHIMURA und P. TODD (1997): "Matching as an econometric evaluation estimator: evidence from evaluating a job training program", *Review of Economic Studies*, 64, 605-654.
- HECKMAN, J., H. ICHIMURA, J. SMITH und P. TODD (1998): "Characterizing Selection Bias Using Experimental Data", *Econometrica*, 66, 1017-1098.
- HECKMAN, J., R. LALONDE und J. SMITH (1999): "The Economics and Econometrics of Active Labor Market Programs", in: O. Ashenfelter und D. Card, Hrsg.: *Handbook of Labor Economics*, Vol. III, Amsterdam: Elsevier Science B.V., 1865-2097.
- HECKMAN, J. und J. SMITH (1999): "The Pre-Program Earnings Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies", *Economic Journal*, 109, 313-348.
- HENGARTNER, N., M. WEGKAMP, und E. MATZNER-LOBER (2002): "Bandwidth selection for local linear regression smoothers", *Journal of the Royal Statistical Society, Ser. B*, 64, 791-804.
- HÜBLER, O. (2001): "Evaluation of policy interventions: Measurement and problems", *Allgemeines Statistisches Archiv*, 85, 103-126
- HUJER, R. und M. CALIENDO (2000): "Evaluation of Active Labour Market Policy: Methodological Concepts and Empirical Estimates", *IZA Discussion Paper No. 236*, Forschungsinstitut zur Zukunft der Arbeit, Bonn.
- HUJER, R., M. CALIENDO und D. RADIĆ (2001), "Nobody Knows... How Do Different Evaluation Estimators Perform in a Simulated Labour Market Experiment?", *Discussion Paper*, <http://much-magic.wiwi.uni-frankfurt.de/> Professoren/hujer/ (2002-5-13).
- HUJER, R., K.-O. MAURER und M. WELLNER (1999): "The Effects of Public Sector Sponsored Training on Unemployment Duration in West Germany – A Discrete Hazard Rate Model based on a Matched Sample", *ifo-Studien*, 45, 371-410.
- HUJER, R. und M. WELLNER (2000): "The Effects of Public Sector Sponsored Training on Individual Employment Performance in East Germany", *IZA Discussion Paper No. 141*, Forschungsinstitut zur Zukunft der Arbeit, Bonn.
- KAUFMANN, H. und PAPE, H. (1996): "Clusteranalyse", in: Fahrmeir, L., A. Hamerle und G. Tutz, Hrsg.: *Multivariate statistische Verfahren*, 2. Aufl., Berlin: de Gruyter, 437-536.

- LALONDE, R. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *American Economic Review*, 76, 604-620.
- LARSSON, L. (2000), "Evaluation of Swedish Youth Labour Market Programmes", *Scandinavian Working Papers in Economics No. 2000:1*, Uppsala University.
- LECHNER, M. (2002): "Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods", *Journal of the Royal Statistical Society*, Ser. A, 165/1, 59-82.
- LECHNER, M. (2001): "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption", in: Lechner, M. und F. Pfeiffer, Hrsg: *Econometric Evaluation of Labour Market Policy*, Heidelberg: Physica/Springer, 43-58.
- LECHNER, M. (1998): "Training the East German Labour Force. Microeconomic Evaluations of Continuous Vocational Training after Unification", Heidelberg: Physica.
- OPITZ, O. (1980): "Numerische Taxonomie", Stuttgart, New York: Fischer.
- REINOWSKI, E., B. SCHULTZ und J. WIEMERS (2003): „Evaluation von Maßnahmen der aktiven Arbeitsmarktpolitik mit Hilfe eines iterativen Matching-Algorithmus. Eine Fallstudie über langzeitarbeitslose Maßnahmeteilnehmer in Sachsen“, *IWH Discussion Paper No. 173*, Institut für Wirtschaftsforschung Halle, Halle.
- REINOWSKI, E., B. SCHULTZ und J. WIEMERS (2004): "Evaluation of Further Training Programmes with an Optimal Matching Algorithm", *IWH Discussion Papers No. 188*, Institut für Wirtschaftsforschung Halle, Halle.
- ROSENBAUM, P. (1989): "Optimal Matching for Observational Studies", *Journal of the American Statistical Association*, 84, 1024-1032.
- ROSENBAUM, P. (1991): „A Characterization of Optimal Designs for Observational Studies“, *Journal of the Royal Statistical Society*, Ser. B, 53, 597-610.
- ROSENBAUM, P. (2002): "Observational Studies", 2. Aufl. New York: Springer.
- ROSENBAUM, P. und D. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70, 41-55.
- ROSENBAUM, P. und D. RUBIN (1985): "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score", *The American Statistician*, 39, 33-39.
- SCHMIDT, C. (1999): "Knowing what Works – the Case for Rigorous Program Evaluation", *IZA Discussion Paper No. 77*, Forschungsinstitut zur Zukunft der Arbeit, Bonn.
- SEIFERT, B. UND T. GASSER (2000): "Data Adaptive Ridging in Local Polynomial Regression", *Journal of Computational and Graphical Statistics*, 9., 338-360.