

Mikroökonometrische Evaluation arbeitsmarktpolitischer Massnahmen

Michael Lechner

September 2002 Discussion paper no. 2002-20

Editor: Prof. Jörg Baumberger
University of St. Gallen
Department of Economics
Bodanstr. 1
CH-9000 St. Gallen
Phone ++41 71 224 22 41
Fax ++41 71 224 28 85
Email joerg.baumberger@unisg.ch

Publisher: Forschungsgemeinschaft für Nationalökonomie
an der Universität St. Gallen
Dufourstrasse 48
CH-9000 St. Gallen
Phone ++41 71 224 23 00
Fax ++41 71 224 26 46

Electronic Publication: www.fgn.unisg.ch/public/public.htm

Mikroökonometrische Evaluation arbeitsmarktpolitischer Massnahmen

Michael Lechner¹

Author's address:

Prof. Dr. Michael Lechner
Schweizerisches Institut für Aussenwirtschaft und Angewandte
Wirtschaftsforschung
Dufourstrasse 48
9000 St. Gallen
Tel. ++41 71 224 23 40
Fax ++41 71 224 22 98
Email Michael.Lechner@unisg.ch
Website www.siaaw.unisg.ch

¹Ich bedanke mich beim Schweizerischen Nationalfond (Projekte 4045-058311, 4045-050673) für finanzielle Unterstützung und bei Wolfgang Franz und Roman Liesenfeld für hilfreiche Kommentare.

Abstract

Many European countries entertain considerable active labour market policies (ALMP) to increase the reemployment chances of the unemployed. Since policies are very costly the pressure for scientific evaluations of the single programmes that are part of the ALMP increases. This paper surveys recent developments in the methodology of microeconomic evaluation that emphasises the issue of nonparametric identification of causal effects as a pre-requisite for reliable estimation and policy conclusions. The paper concludes with some brief remarks about recent empirical evaluations of the Swiss ALMP.

Zusammenfassung

Viele europäische Länder verwenden neben einer passiven Einkommensabsicherung der Arbeitslosen die aktive Arbeitsmarktpolitik, um deren Wiederbeschäftigung zu beschleunigen. Zunehmend setzt sich dabei die Erkenntnis durch, dass diese kostenintensive Politik dem Steuerzahler nur bei nachweisbarem Nutzen langfristig zu vermitteln ist. Auch wenn die Bereitstellung der für seriöse Evaluationen notwendigen Daten in Europa nur langsam vorankommt, so gab es auf der methodischen Seite der mikroökonomischen Evaluationsstudien erhebliche Weiterentwicklungen. Diese methodischen Entwicklungen stehen im Mittelpunkt dieser Arbeit. Die mikroökonomische Evaluationsliteratur hat die Frage nach den Bedingungen für eine nichtparametrische Identifikation heterogener kausaler Effekte in den Vordergrund gestellt. Damit wurde auch althergebrachten Schätzverfahren zu neuer Bedeutung und zu einer neuen Interpretation verholfen. Es ist anzunehmen, dass sich das in dieser Literatur entwickelte Denkmuster auch im Mainstream der angewandten (Mikro- und Panel-) Ökonometrie der Zukunft etablieren wird, denn immerhin es das Ziel eines jeden angewandten ökonometrischen Papiers, kausale Schlussfolgerungen zu ziehen. Ausserdem wird in dieser Arbeit auf die Datenproblematik und die wichtigsten empirischen Ergebnisse für die Schweiz kurz eingegangen.

Keywords

Aktive Arbeitsmarktpolitik, Identifikation kausaler Effekte, Nichtparametrische Schätzung

JEL Classification

J68

1 Einleitung

Heutzutage betreiben viele europäische Länder neben einer passiven Einkommensabsicherung der Arbeitslosen eine substantielle aktive Arbeitsmarktpolitik (AAMP) mit Fortbildungskursen, Beschäftigungsprogrammen, Lohnsubventionen, etc., um deren Wiederbeschäftigung zu beschleunigen. Zunehmend setzt sich jedoch die Erkenntnis durch, dass eine derartig kostenintensive Politik dem Steuerzahler nur bei (mehr oder weniger) wissenschaftlich nachweisbarem Nutzen langfristig zu vermitteln ist.

In den letzten Jahren waren die grössten Anstrengungen in diese Richtung in Schweden, Grossbritannien der Schweiz, aber auch in anderen kleineren europäischen Ländern zu beobachten. In Schweden wurde eigens ein staatliches Wirtschaftsforschungsinstitut gegründet (IFAU, angegliedert an die Universität Uppsala), dessen einzige Aufgabe es ist, die schwedische Arbeitsmarktpolitik, und insbesondere die AAMP, mit Mikro- und Makrodaten auf wissenschaftlichem Niveau zu evaluieren. In Grossbritannien ist es ein explizites Ziel der Regierung die Wirkung der einzelnen Komponenten der AAMP, die dort im wesentlichen unter dem Etikett *New Deal for ...* firmieren, regelmässig zu überprüfen und die Politik entsprechend anzupassen.¹ Praktisch geschieht dies durch ständige Ausschreibungsrunden und Bieterwettbewerbe der Regierung, die damit immer weitere Wirtschaftsforschungsinstitute und Beratungsfirmen dazu bringt, sich in diesem Bereich zu engagieren.² In der Schweiz gab es 1998 die erste grosse Evaluationsrunde. Wirtschaftsforschungsinstitute und Forschergruppen an den Universitäten wurden, eingeladen die Wirkungen der AAMP zu untersuchen. Inzwischen ist eine weitere Runde geplant.

Auch in Deutschland ist ein gewisses Interesse an einer ernstzunehmenden Evaluation der AAMP entstanden. Es sind vermehrte Anstrengungen zu beobachten eine Datensituation zu erreichen, die es erlauben wird, die im wissenschaftlichen Diskurs entwickelten Methoden und Verfahren für eine Verbesserung der

¹ Siehe Dolton (2002) für einen Überblick.

² Diese Anstrengungen werden in der Zukunft manchmal zynischerweise als *New Deal for Econometricians* bezeichnet.

wirtschaftspolitischen Praxis im Bereich der AAMP einzusetzen.³ Jedoch wird bis zum Vorliegen erster Studien auf Basis der neuen Daten noch einige Zeit vergehen.

Auch wenn die Bereitstellung der für seriöse Evaluationen notwendigen Daten zumindest in Europa lange Zeit nur langsam vorankam, so gab es auf der methodischen Seite der mikroökonomischen Evaluationsstudien erhebliche Weiterentwicklungen, die in den Übersichtsarbeiten von Angrist, Krueger (1999), Heckman, LaLonde, Smith (1999) und Heckman, Vytlačil (2002) ausführlich dokumentiert sind.⁴ Dabei wird insbesondere darauf Wert gelegt, die tatsächlichen Effekte der Programme von reinen Selektionseffekten zu trennen, um Scheineffekte, die von der unterschiedlichen Zusammensetzung der Programmteilnehmer herrühren, zu vermeiden.⁵ Die neuesten methodischen Entwicklungen versuchen die Vielfältigkeit typischer europäischer Programme explizit zu berücksichtigen. Diese Programme zeichnen sich zum einen durch eine Vielzahl unterschiedlicher Programmtypen aus, und zum anderen durch Sequenzen einzelner Programme, die der Arbeitslose nacheinander durchläuft.⁶ Die Ergebnisse dieser Literatur lassen sich auch verwenden, die Allokation von Teilnehmern auf einzelne Massnahmen zu verbessern.⁷

³ Eine über das Thema AAMP weit hinausgehende Beurteilung der Datensituation in Deutschland enthält das Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten KOMMISSION ZUR VERBESSERUNG DER INFORMATIONELLEN INFRASTRUKTUR ZWISCHEN WISSENSCHAFT UND STATISTIK (2001). Neben einer Bestandsaufnahme enthält das Gutachten Empfehlungen, die zu einer verbesserten Datenverfügbarkeit in Deutschland führen würden.

⁴ Siehe auch die Beiträge in Lechner und Pfeiffer (2001).

⁵ Beispiel: Werden in ein Programm systematisch Personen geschickt, die sowieso recht leicht einen Arbeitsplatz finden würden, und in ein anderes Programm im wesentlichen Personen mit grossen Problemen am Arbeitsmarkt, so werden die Teilnehmer des ersten Programms im Schnitt immer leichter einen Arbeitsplatz finden, als die Teilnehmer des zweiten Programms. Der Unterschied der Erwerbschancen auf dem Arbeitsmarkt ist dann die Summe des Selektionseffekts (ungleiche Teilnehmerzusammensetzung), der augenscheinlich nichts oder nur wenig mit der Effektivität des Programms zu tun haben muss, und des Programmeffekts, um den es eigentlich geht. Die Isolation der einzelnen Effekte ist das zentrale Problem. James Heckman erwarb sich in diesem Bereich grosse Verdienste, die durch die Verleihung des Nobelpreises in Ökonomie im Jahr 2000 gewürdigt wurden.

⁶ Siehe Imbens (2000) und Lechner (2001a, 2002b, c) für heterogene Programme und Lechner und Miquel (2001) und Miquel (2002a, b) für Programmsequenzen.

⁷ Siehe z.B. Frölich (2002).

Da es einerseits schon umfangreiche Übersichtsarbeiten zu den existierenden empirischen Evaluationsstudien gibt, und andererseits für viele Länder die Ergebnisse der empirischen Studien auf einer oft (noch?) zweifelhaften Datenbasis beruhen, wird in diesem Aufsatz der methodische Teil der mikroökonomischen Evaluationsliteratur im Mittelpunkt stehen. Er wird im nächsten Abschnitt 2 (selektiv) dargestellt werden. Abschnitt 3 diskutiert kurz die Datenproblematik. Abschnitt 4 enthält neben einer Übersicht über publizierte Übersichtsartikel eine knappe Darstellung der wichtigsten empirischen Ergebnisse für die Schweiz, die auf einer relativ soliden Datenbasis beruhen. In Abschnitt 5 werden einige Schlussfolgerungen gezogen.

2 Die Prinzipien der modernen mikroökonomischen Evaluationsforschung

Das Vorgehen angewandter ökonomischer Studien ist immer das gleiche: Mit Hilfe mehr oder minder ausgefeilter statistischer Methoden werden Grössen bedingter Verteilungen (Koeffizienten, Momente, etc.) geschätzt und diese dann kausal interpretiert. Je nach Mut des Autors und Zielrichtung der Studie werden im Anschluss mehr oder weniger präzise Politikempfehlungen gegeben. Nach wie vor wesentlich seltener findet man jedoch, dass die Autoren sehr viel Zeit und Energie darauf verwenden, ihre kausale Fragestellung gezielt zu modellieren und sich über die Bedingungen Sorgen zu machen, die erfüllt sein müssen, um von empirischen Zusammenhängen der Variablen X und Y auf deren kausale Beziehungen schliessen zu können. Oder um es mit einem Satz zu sagen, der dem Nobelpreisträger James Heckman zugeschrieben wird: *We always estimate something. The hard part is to find out what it means.* Mit anderen Worten, die Frage lautet, unter welchen Annahmen man den geschätzten Werten, die üblicherweise in engem Zusammenhang zu partiellen Korrelationen stehen, eine kausale Interpretation geben kann. Diese Annahmen werden auf externem Wissen basieren müssen, das mit den Daten nicht überprüfbar ist.

An dieser Stelle setzt das methodische Vorgehen der modernen Evaluationsforschung an: Zuerst wird die kausale Fragestellung explizit dargestellt und formalisiert. Daran anschliessend folgen intensive Überlegungen, unter welchen Annahmen und mit welchen aus den Daten gewinnbaren Grössen auf diese kausale

len Zusammenhänge zurückgeschlossen werden könnte (Identifikation). Erscheinen die identifizierenden Annahmen plausibel, werden die interessierenden Effekte geschätzt. Sofern die Stichprobe gross genug ist, wird man Schätzverfahren vorziehen, die im wesentlichen nur die identifizierenden Annahmen verlangen. Abschliessend wird versucht die Ergebnisse wirtschaftspolitisch zu interpretieren.

2.1 Die Definition der zu beantwortenden Frage

Um Aussagen über die Wirkung einer arbeitsmarktpolitischen Massnahme auf ein einzelnes Individuum treffen zu können, ist es notwendig, formal korrekt zu beschreiben, was man unter einer (kausalen) Wirkung versteht. Insbesondere gilt es, den kausalen Effekt einer Massnahmenteilnahme von einer Korrelation der Massnahmenteilnahme mit einer Ergebnisvariable (z. B. Arbeitsmarktstatus oder Einkommen 2 Jahre nach der Massnahme) zu unterscheiden.⁸ Für ökonometrische Evaluationen hat sich weitgehend das Modell potentieller Ergebnisse (Neyman, 1923, 1935, Roy, 1951, Rubin, 1974) durchgesetzt. Um den formalen Apparat gering zu halten, wird im folgenden vom einfachsten Prototyp dieses Modells (statisch, nur 2 Zustände) ausgegangen. Am Ende des Abschnitt erfolgen Hinweise auf Erweiterungen.

Die Grundidee des Roy-Rubin-Modells besteht in einem Vergleich der (potentiellen) Ergebnisse des Individuums bei einer Massnahmenteilnahme (Y^1) mit dem Ergebnis bei Nichtteilnahme (Y^0). Als kausaler Effekt wird üblicherweise die Differenz der potentiellen Ergebnisse ($Y^1 - Y^0$) aufgefasst. Entscheidend ist also der Vergleich zweier Zustände, von denen nur ein Zustand beobachtbar und der andere Zustand unbeobachtbar ist. Damit man hier von Kausalität sprechen kann, ist die klassische Partialannahme notwendig, nämlich die Unabhängigkeit der individuellen potentiellen Ergebnisse von dem Teilnahmeverhalten anderer Individuen (*stable unit treatment value*-Annahme, SUTVA). Der jeweilige Zustand des Individuums wird durch die binäre Zufallsvariable S ($S=1$ bedeutet Teilnahme) beschrieben, deren Realisationen beobachtet werden können. Die beobachtbare Ergebnisvariable Y ist naturgemäss eine Funktion von S und den eigentlich interessierenden potentiellen Ergebnissen:

$$Y = Y^0(1 - S) + Y^1S. \quad (1)$$

Da für jedes Individuum eines der beiden Ereignisse unbeobachtbar sein muss, ist es offensichtlich unmöglich (ausser mit vollkommen unplausiblen Annahmen), einen individuellen kausalen Effekt abzuschätzen. Hat man jedoch eine grössere Stichprobe mit Teilnehmern und Nichtteilnehmern an der Massnahme zur Verfügung, so kann es möglich sein, bestimmte durchschnittliche kausale Effekte zu identifizieren (siehe Abschnitt 2.2), von denen die drei populärsten Vertreter in den folgenden Gleichungen definiert werden:

$$\gamma_0 := E(Y^1 - Y^0) = E(Y^1) - E(Y^0), \quad (2)$$

$$\theta_0 := E(Y^1 - Y^0 \mid S = 1) = E(Y^1 \mid S = 1) - E(Y^0 \mid S = 1), \quad (3)$$

$$\xi_0 := E(Y^1 - Y^0 \mid S = 0) = E(Y^1 \mid S = 0) - E(Y^0 \mid S = 0). \quad (4)$$

γ_0 steht für den erwarteten Massnahmeneffekt eines zufällig aus der interessierenden Population ausgewählten Individuums (*average treatment effect*, ATE).⁹ Entsprechend stehen θ_0 und ξ_0 jeweils für den Effekt eines aus der Gruppe der Teilnehmer (*average treatment effect on the treated*, ATET) bzw. aus der Gruppe der Nichtteilnehmer ausgewählten Individuums. In vielen Evaluationsstudien steht der Parameter θ_0 im Vordergrund, denn θ_0 gibt den Massnahmenerfolg für Personen an, die tatsächlich am Programm teilgenommen haben und kann somit als ein Erfolgsindikator des Programm aufgefasst werden. ξ_0 wäre relevant, wenn man aus der Gruppe der Nichtteilnehmer weitere Personen ins Programm aufnehmen würde. Wenn die Selektion effizient in dem Sinne ist, dass Personen mit höheren Programmwirkungen mit

⁸ Siehe z.B. Cox (1992), Dawid (200), Holland (1986, 1989), Leamer (1983), Pearl (2000), Leamer, Pratt, Schlaifer (1988) und Sobel (1994) für ausführliche Diskussionen der Kausalitätskonzepte in Ökonometrie, Statistik und anderen Wissenschaften.

⁹ Die folgenden Überlegungen sind in gleicher Form auch in Teilpopulationen gültig, sofern diese durch Variablen definiert werden, die nicht durch S beeinflusst sind.

grösserer Wahrscheinlichkeit am Programm teilnehmen, dann implizierte dies $\theta_0 > \xi_0$. Keiner der definierten Effekte verlangt a priori, dass der Effekt von S auf Y homogen für alle Individuen ist, ganz im Gegenteil wird individuelle Heterogenität (nahezu) uneingeschränkt zugelassen.

Oft wäre es wünschenswert, mehr über die Verteilung der individuellen kausalen Effekte zu erfahren als nur den Erwartungswert. Die meisten anderen statistischen Masse für die Verteilung der kausalen Effekte (Varianz, Median, etc.) haben jedoch den aus ökonometrischer Sicht gravierenden Nachteil, dass sie von Eigenschaften der gemeinsamen Verteilung von (Y^1, Y^0) abhängen und nicht allein anhand der marginalen Verteilungen von Y^1 und Y^0 identifizierbar sind, wie dies beim Erwartungswert der Fall ist. Während jedoch Realisationen aus den marginalen Verteilungen $f(Y^1 | S = 1)$ ($= f(Y | S = 1)$) und $f(Y^0 | S = 0)$ ($= f(Y | S = 0)$) in der Regel zur Verfügung stehen, sind gleichzeitige Realisationen von Y^1 und Y^0 nie beobachtbar. Entsprechend strenger müssten die für die Identifikation notwendigen zusätzlichen Annahmen sein.¹⁰ Unter gewissen Annahmen¹¹ kann man die Differenz der Quantile der marginalen Verteilungen als Quantil der Differenz auffassen und entsprechende Quantileffekte definieren (siehe Abadie, Angrist, Imbens, 2000, und Chernozhukov, Hansen, 2001).

Das prototypische mikroökonometrische Modell zur Erfassung kausaler Beziehungen wurde in zwei wesentliche Richtungen erweitert: Imbens (2000) und Lechner (2001a) analysieren den Fall einer endlichen Anzahl diskreter Ereignisse ($S=0,...,M$). Sie zeigen, dass es konzeptionell keine grundlegenden Unterschiede zum binären Fall gibt. Die von Robins (1986), Miquel, Lechner (2001), Miquel (2002a, b) und Lechner (2002d) durchgeführten Arbeiten für dynamische Ereignisse bringen hingegen konzeptionelle Erweiterungen, insbesondere hinsichtlich Definition und Notwendigkeit von Exogenitätsannahmen.

¹⁰ Heckman, Smith, Clements (1997) zeigen, dass einige derartige Masse im Falle eines sozialen Experiments auf ein Intervall eingeschränkt werden können. Die Intervallbreite ist allerdings oft sogar zu gross, um bezüglich des Vorzeichens eindeutige Ergebnisse zu erhalten. Siehe auch die Diskussion in Heckman, Lalonde, Smith (1999).

¹¹ Y^1 und Y^0 müssen die gleiche Ordnung aufweisen, d.h. wenn das Einkommen einer Person bei potentieller Teilnahme am 1. Dezil liegt, dann muss es auch bei potentieller Nichtteilnahme am 1. Dezil liegen usw.

2.2 Das Identifikationsproblem

Ist man an einem in den Gleichungen (2) bis (4) definierten kausalen Effekte interessiert, so wird klar, dass man diesen Effekt ohne weitere Annahmen nicht konkret abschätzen können. Das Problem besteht darin, dass die auf beobachtbaren Größen basierenden bedingten Erwartungswerte $E(Y | S = 1)$ und $E(Y | S = 0)$ oft nicht den entsprechenden Erwartungswerten der potentiellen Ergebnisse $E(Y^1 | S = 0)$ und $E(Y^0 | S = 1)$ entsprechen. Vergleicht man beispielsweise die mittleren Erwerbseinkommen der Teilnehmer einer Massnahme ($E(Y | S = 1)$), in der überwiegend Personen mit schlechten Arbeitsmarktchancen zugewiesen werden, so wäre es nicht verwunderlich, wenn diese Personen im Vergleich zu Nichtteilnehmern ($E(Y | S = 0)$) erheblich weniger verdienen, da dieser Vergleich sowohl den Effekt der Massnahme, als auch den Effekt der unterschiedlichen Teilnehmerstrukturen enthält:¹²

$$\begin{aligned} E(Y | S = 1) - E(Y | S = 0) &= E(Y^1 | S = 1) - E(Y^0 | S = 0) = \\ &= \underbrace{E(Y^1 | S = 1) - E(Y^0 | S = 1)}_{\text{Kausaler Effekt von } S} + \underbrace{E(Y^0 | S = 1) - E(Y^0 | S = 0)}_{\text{Selektionseffekt}}. \end{aligned} \quad (5)$$

Im nächsten Unterabschnitt werden einige in der Literatur geläufige identifizierende Annahmen und darauf abgestimmte Schätzverfahren vorgestellt. Nur wenn diese Annahmen plausibel sind – und darüber lässt sich oft trefflich streiten – sollte man die Schätzungen kausal interpretieren. Diese empirisch nicht testbaren Annahmen benötigen immer externes Wissen über den Zusammenhang zwischen Selektionsmechanismus und Ergebnisvariablen. Generell lässt sich aber festhalten, dass je informativer der zur Verfügung stehende Datensatz ist, desto leichter sind glaubwürdige Annahmen zu finden.¹³

¹² Im Laufe dieses Abschnitts sollte es deutlich werden, dass die Konzepte der Selektivität, Kausalität und Endogenität unterschiedliche Seiten der gleichen Pyramide darstellen.

¹³ Welche Informationen in welcher Art ausgenutzt werden können wird im folgenden erläutert.

Im Falle arbeitsmarktpolitischer Massnahmen benötigt man neben ökonometrischen Kenntnissen auch Kenntnisse über das Entscheidungsverhalten der einzelnen Akteure (Arbeitslose und Arbeitsvermittler) und die Art und Weise, wie die jeweils zur Verfügung stehenden Daten zustande gekommen sind.

Evaluationen sind kontextbezogen. Sie müssen auf die Daten, das jeweilige institutionelle Umfeld und das Entscheidungsverhalten der relevanten ökonomischen Agenten abgestimmt sein. Manchmal verlangt ein bestimmter Kontext sogar explizite ökonometrische Weiterentwicklungen, um dessen Spezifika adäquat Rechnung tragen zu können. Es gibt generell kein bestes Evaluationsverfahren, sondern es gibt nur das Verfahren, das für eine spezifische Evaluationssituation am besten zu passen scheint. Diese Problematik ist nichts Neues und wird treffend im folgenden Zitat von Heckman, Robb (1986) dargestellt:

*Social scientists never have access to true experimental data of the type sometimes available to laboratory scientists. Our inability to use laboratory methods to independently vary treatments to eliminate or isolate spurious channels of causation places a fundamental limitation on the possibility of objective knowledge in the social sciences. In place of laboratory experimental variation, social scientists use subjective thought experiments. Assumptions replace data. In the jargon of modern econometrics, minimal identifying assumptions are invoked. Because minimal identifying assumptions cannot be tested with data (all possible minimal assumptions for a model explain the observed data equally well, at least in large samples) and because empirical estimates of causal relationships are sensitive to these assumptions, inevitably there is scope for disagreement in the causal interpretation of social science data. Context, beliefs, and a priori theory resolve differences in causal interpretations. By definition, the available data cannot do so, although in principle, experiments can. The solution to the problem of causal inference lies outside of mathematical statistics and depends on contexts which are not universal."*¹⁴

2.3 Einige identifizierende Annahmen und dazu korrespondierende Schätzverfahren

Im folgenden werden unterschiedliche identifizierende Annahmen und dazu korrespondierende Schätzverfahren diskutiert. Die Reihenfolge in diesem Abschnitt ist daran orientiert, inwieweit die zur Verfügung stehenden Daten alle notwendigen Informationen über den relevanten Selektionsmechanismus enthalten. Die Darstellung ist weitgehend knapp und informell gehalten.¹⁵

¹⁴ Heckman and Robb (1986), S. 63.

¹⁵ Für Details siehe die exzellenten Übersichtsarbeiten von Angrist, Krueger (1999), Heckmann, Lalonde, Smith (1999), Heckman, Robb (1985) und Heckman, Vytlacil (2002).

2.3.1 Das ideale soziale Experiment

Wählt man aus den Anmeldungen für eine arbeitsmarktpolitische Massnahme zufällig eine Gruppe aus, denen man den Zugang zu dieser Massnahme verweigert, und vergleicht die Mittelwerte der Ergebnisvariable (Y) für beide Gruppen, so besitzt man eine einfache und überzeugende Schätzung für einen mittleren kausalen Effekt für die Teilnehmer (der Effekt für die Nichtteilnehmer und der Population ist in einem einfachen experimentellen Design nicht identifiziert), denn die zufällige Auswahl im Experiment eliminiert den Selektionseffekt für ATET ($E(Y^0 | S = 1) = E(Y^0 | S = 0)$), da S unabhängig von Y^0 durch den Computer realisiert wurde).

Soziale Experimente, die grösstenteils in den USA durchgeführt wurden,¹⁶ aber inzwischen auch in Deutschland zu finden sind,¹⁷ weisen auch bei einer idealen Durchführung Nachteile auf. Zum einen sind sie teuer und es vergeht viel Zeit vom Beginn des Experiments bis zum Vorliegen der Evaluationsergebnisse. Zum anderen gibt es zumindest in Europa oft ethische oder sonstige Vorbehalte dagegen, gewisse vermutete Vorteile einer neuen Massnahme einer zufällig ausgewählten Gruppe vorzuenthalten. Hinzu kommen die Probleme der sogenannten Substitutionsverzerrung (Nichtteilnehmer nehmen z. B. an einem ähnlichen Kurs ausserhalb des Programms teil) und der Aussteigeverzerrung (ausgewählte Teilnehmer verweigern die Teilnahme, bzw. brechen den Kurs ab). Das Pro und Contra experimenteller Evaluationsstudien wird ausführlich in Burtless (1995) und Heckman, Smith (1995) diskutiert.

In der deutschen Arbeitsmarktpolitik könnte man sich vorstellen, dass Experimente in einer marginalen Variante in Zukunft eine wichtige Rolle spielen werden. Um der Akzeptanz willen wird zu beachten sein, dass keine Leistungen, die von den Betroffenen als sehr wichtig eingestuft werden, tangiert werden. Aber man könnte z.B. sehr gewinnbringend die Durchführung von Beratungsaktivitäten und Programmen experimentell variieren, um so deren Implementation zu optimieren.

¹⁶ Siehe die Übersicht in Greenberg, Shroder (1997).

¹⁷ Siehe die Arbeit von Rudolph, Müntnich (2001), die den Profilingansatz für Arbeitslose experimentell beurteilen.

2.3.2 Der ideale nichtexperimentelle Datensatz und die CIA

Die Verzerrung $[E(Y^0 | S = 1) - E(Y^0 | S = 0)]$ wird durch Variablen hervorgerufen, die sowohl mit der Selektion der Teilnehmer als auch mit den potentiellen Ergebnisvariablen korreliert sind (z.B. das Bildungsniveau). Sind diese Variablen (X) bekannt und deren Ausprägungen im Datensatz beobachtbar, kann das Selektionsproblem gelöst werden, indem man für diese Variablen in einer Schätzung *kontrolliert*: Für einen festen Wert dieser Variablen gibt es definitionsgemäss keine Selektionsverzerrung $[E(Y^0 | X = x, S = 1) = E(Y^0 | X = x, S = 0) = E(Y | X = x, S = 0)]$ (Rubin, 1977). Da im Falle der Gültigkeit dieser Annahme S und Y^0 für jeden Wert von X unabhängig sind, spricht man auch von der Annahme bedingter Unabhängigkeit (*conditional independence assumption*, CIA). Der resultierende Ausdruck für ATET ist konsistent schätzbar, da er nur von beobachtbaren Grössen abhängt:

$$\begin{aligned}\theta_0 &= E(Y^1 | S = 1) - E(Y^0 | S = 1) = E(Y | S = 1) - \underset{X|S=1}{E}[E(Y | X = x, S = 0)] \\ &= E(Y | S = 1) - \underset{p(x)|S=1}{E}[E(Y | p(X) = p(x), S = 0)]; \\ p(x) &:= P(S = 1 | X = x); \quad p(x) < 1.\end{aligned}\tag{6}$$

Man vergleicht die Teilnehmer mit Nichtteilnehmern, die gleiche Charakteristika X aufweisen (die wiederum nicht von S beeinflusst werden dürfen; siehe z.B. Rosenbaum, 1984). Rosenbaum und Rubin (1983) zeigen, dass die Vergleichbarkeit bzgl. X ein unnötig strenges Kriterium darstellt: Ein Vergleich von Teilnehmern mit Nichtteilnehmern mit der gleichen bedingten Teilnahmewahrscheinlichkeit (dem sogenannten *propensity score*) reicht aus.

Die Glaubwürdigkeit dieser (datenhungrigen) Identifikationsstrategie CIA steht und fällt damit, ob man zum einen glaubwürdig für das zugrundeliegenden institutionelle Umfeld darlegen kann, welche Variablen plausiblerweise in X enthalten sein müssten, und ob diese Variablen im für die empirische Studie vorhandenen Datensatz zumindest annähernd enthalten sind. Sind diese Punkte nicht erfüllt, erübrigt sich jede weitere Diskussion über mögliche Schätzverfahren, die auf CIA basieren.

Die Schätzung des kritischen zweiten Terms in Gleichung (6) kann erfolgen, indem in einem ersten Schritt $E(Y | X = x, S = 0)$ oder $E(Y | p(X) = p(x), S = 0)$ mit Hilfe einer parametrischen oder vorzugsweise nichtparametrischen Regression geschätzt wird. Anschliessend wird diese Funktion hinsichtlich der Verteilung der Charakteristika in der Gruppe der Teilnehmer gewichtet gemittelt. Gibt es ein plausibles Modell für die bedingte Teilnahmewahrscheinlichkeit, so ist die Schätzung von $E(Y | p(X) = p(x), S = 0)$ anstatt von $E(Y | X = x, S = 0)$ attraktiv, da der sogenannten *Fluch der Dimensionalität* bei nichtparametrischen Regressionen vermieden werden kann (siehe z. B. Frölich, 2002a). Parametrische Regressionen (OLS, Tobit, usw.) verwenden überidentifizierende Restriktionen und sind eigentlich nur in kleinen Stichproben interessant, in denen sich nichtparametrische Methoden erfahrungsgemäss sehr schwer tun.

In empirischen Studien kommt sehr oft ein anderes, pseudo-experimentelles Verfahren zur Anwendung, das explizit auf den Zwischenschritt der Regression verzichtet und direkt $E_{X|S=1} [E(Y | X = x, S = 0)]$, bzw.

$E_{p(x)|S=1} [E(Y | p(X) = p(x), S = 0)]$ schätzt. Die Idee dieser sogenannten Matchingverfahren ist einfach:

Finde für jeden Teilnehmer einen Nichtteilnehmer, der diesem entweder bzgl. X oder $p(X)$ sehr ähnlich ist. Der Mittelwert dieser Vergleichsgruppe ist das gesuchte Schätzverfahren. Matchingverfahren wurden von Rubin (1979) formal eingeführt, lassen sich aber bis auf die Arbeiten von Fechner (1860) zurückführen. Matchingverfahren gibt es in verschiedenen Varianten, deren Diskussion den Rahmen dieses Übersichtsartikels sprengen würde.¹⁸

Da die empirischen Anwendungen mit Hilfe dieser Methoden inzwischen so umfangreich sind, dass sich eine vollständige Aufzählung verbietet, nenne ich nur einige Evaluationsstudien für Deutschland und die Schweiz: Für Deutschland sind dies beispielsweise Fitzenberger, Speckesser (2002), Hujer, Maurer, Wellner (1997), Hujer, Wellner (1998), und Lechner (1998). Für die Schweiz wäre die Arbeit von Gerfin,

¹⁸ Siehe Dehejia, Wahba (1999, 2002), Heckman, Smith (1999), Heckman, Robb (1986), Heckman, Ichimura, Todd (1997), Heckman, Ichimura, Smith, Todd (1998), Rosenbaum, Rubin (1985), Rosenbaum, (1995), Smith, Todd (2002), und Lechner (1999a, 2002b, c).

Lechner (2002) zu nennen.¹⁹ Die Schweiz ist ein interessanter Fall, da dort erstmals auf Wunsch der Evaluatoren auf Basis verknüpfter Registerdaten ein informativer Datensatz für die Durchführung einer Evaluation der AAMP mit Hilfe der CIA zusammengestellt wurde. Diesen Weg haben inzwischen auch die Bundesanstalt für Arbeit und einige interessierte deutsche Wirtschaftsforschungsinstitute eingeschlagen.

2.3.3 Im Datensatz fehlen wichtige Variablen I: Instrumentalvariablen und Kontrollfunktionen

Erscheint CIA nicht plausibel, so sind andere (weitere) Informationen notwendig, um kausale Wirkungsaussagen treffen zu können. Ein zentraler Ansatz, um mit der durch die fehlenden Variablen ausgelösten Endogenität von S umzugehen, ist der altbekannte Instrumentalvariablenansatz, der für den Fall heterogener kausaler Effekte durch die Arbeiten von Imbens, Angrist (1994), Angrist, Imbens, Rubin (1996) für diskrete Instrumente und Heckman, Vytlačil (1999) für kontinuierliche Instrumente eine neue Interpretation und Bedeutung erlangte.²⁰ Zur Veranschaulichung wird hier nur der einfachste Fall eines binären Instrumentes Z mit den möglichen Ausprägungen 0 und 1 kurz dargestellt. In diesem Fall entspricht das übliche IV-Schätzverfahren folgendem Ausdruck (Durbin, 1954):

$$\hat{\beta}_N^{IV} = \frac{Cov_N(y_i, z_i)}{Cov_N(s_i, z_i)} = \frac{\sum_{i=1}^N y_i z_i / \sum_{i=1}^N z_i - \sum_{i=1}^N y_i (1 - z_i) / \sum_{i=1}^N (1 - z_i)}{\sum_{i=1}^N s_i z_i / \sum_{i=1}^N z_i - \sum_{i=1}^N s_i (1 - z_i) / \sum_{i=1}^N (1 - z_i)}. \quad (7)$$

Es konvergiert somit unter den üblichen Annahmen gegen folgenden Grenzwert:

$$\beta_0^{IV} = \frac{E(Y | Z = 1) - E(Y | Z = 0)}{P(S = 1 | Z = 1) - P(S = 1 | Z = 0)}. \quad (8)$$

¹⁹ Inzwischen wird dieser Ansatz auch ausserhalb der Evaluationsforschung verwendet. So analysiert Persson (2001) die Wirkung einer Währungsunion, Persson und Tabellini (2001) untersuchen die Wirkungen von Verfassungen und Lechner (2001b) den Geburtenrückgang in den neuen Bundesländern nach der Vereinigung.

²⁰ Angrist, Krueger (2001) ordnen die Instrumentalvariablenverfahren in eine breite, auch methodenhistorische Diskussion zur Suche nach Identifikation ein.

Um β_0^{IV} und somit der üblichen Instrumentalvariablenschätzung eine kausale Interpretation zuweisen zu können, müssen drei wesentliche Annahmen erfüllt sein. Zuerst muss das Instrument mit der endogenen Variablen S korreliert sein (Instrumentenrelevanz, Gleichung (9) unten). Ausserdem darf sein Einfluss auf die potentiellen Ergebnisse *nur* über (den Umweg) S erfolgen ((10), Ausschlussrestriktion).²¹ Da die zweite Annahme nur die potentiellen Ergebnisse betrifft, ist sie nicht testbar.

$$P(S = 1 | Z = z) \neq P(S = 1); \quad (9)$$

$$E[Y^1 | S(Z) = S(z), Z = z] = E[Y^1 | S(Z) = S(z)], \quad E[Y^0 | S(Z) = S(z), Z = z] = E[Y^0 | S(Z) = S(z)]. \quad (10)$$

Für eine kausale Interpretation von β_0^{IV} ist weiterhin eine Annahme notwendig, die besagt, dass alle Individuen auf eine Änderung des Instruments in die gleiche Richtung reagieren, d.h. z.B. wenn der Wert des Instruments steigt, darf es nicht vorkommen, dass einige Individuen mit Teilnahme statt Nichtteilnahme reagieren, während diese Veränderung des Instruments bei anderen Individuen aber zur Nichtteilnahme statt Teilnahme führt ($s_i(1) \geq s_i(0), \forall i$ oder $s_i(1) \leq s_i(0), \forall i$). Für die empirische Analyse besteht üblicherweise die grösste Herausforderung darin, eine Variable zu finden, die mit hoher Plausibilität die Ausschlussrestriktion erfüllt. Sind diese Bedingungen erfüllt, so zeigen die zitierten Arbeiten, dass β_0^{IV} als sogenannter *local average treatment effect* (LATE) kausal interpretierbar ist:

$$\beta_0^{IV} = E[(Y^1 - Y^0) | (|S(1) - S(0)| = 1)]. \quad (11)$$

LATE ist der kausale Effekt von S auf Y für den (unbeobachtbaren) Teil der Bevölkerung, der auf eine Veränderung des Instruments seinen Teilnahmestatus S verändern würde. LATE entspricht im allgemei-

²¹ In der Praxis sind die Annahmen oft erst nach einer weiteren Konditionierung auf exogene Variablen erfüllt. Zur Vereinfachung der Darstellung soll hiervon abgesehen werden. Frölich (2002b) analysiert ausführlich Identifikation und nichtparametrische Schätzung im Falle des Vorhandenseins bzw. der Notwendigkeit von Kovariaten. Er kann zeigen, dass nichtparametrische Schätzverfahren existieren, die $\sqrt{N}$ -konvergent und somit

nen weder ATE noch ATET. Bildet das Instrument eine Politikveränderung ab, so scheint die Interpretation klar, in anderen Fällen mag LATE weniger politikrelevant und somit weniger interessant sein.

Im Gegensatz zum früheren, eher mechanistischen Vorgehen bei der IV-Schätzung, klärt diese neue Literatur weitgehend den Zusammenhang zwischen den notwendigen Bedingungen für die Instrumentengültigkeit und den durch das jeweilige Instrument geschätzten kausalen Effekt. Somit kommt es in der empirischen Anwendung nicht nur darauf an Instrumente zu finden, welche die jeweilige Ausschlussrestriktion erfüllen, sondern die Instrumente sollten auch in bezug auf die Relevanz des durch sie definierten LATE noch eine interessante Variation abbilden. Erhält man bei heterogenen Effekten mit unterschiedlichen Instrumenten verschiedene Ergebnisse, so muss dies nicht Evidenz gegen die Gültigkeit der Instrumente sein, sondern kann vielmehr die unterschiedlichen LATE's widerspiegeln, die von den jeweiligen Instrumenten definiert werden.

Schätzverfahren, die sich bei näherer Untersuchung als den Instrumentalvariablenverfahren sehr ähnlich erweisen, sind Selektionsmodelle und Modelle mit sogenannten *switching regressions*. In diesen Modellen wird die Endogenität (bzw. die resultierende Selektionsverzerrung) von S durch die selektive Beobachtbarkeit dadurch korrigiert, dass der ‚normalen‘ Regression noch eine Kontrollfunktion als weiterer Regressor hinzugefügt wird. Das wahrscheinlich populärste Verfahren dieser Art ist das von Heckman (1979) vorgeschlagene sogenannte Heckit-Verfahren. Die Kontrollfunktion ist diesem Fall die sogenannte Hazardrate, die üblicherweise aus einer Probitschätzung gewonnen wird. Die funktionale Form der Kontrollfunktion wird aus der Annahme gemeinsam normalverteilter Fehlerterme hergeleitet (siehe z.B. Heckman und Robb, 1986) und ist eine nichtlineare Funktion der in den Probit aufgenommenen Variablen. Inzwischen wurden auch semiparametrische Verfahren vorgeschlagen,²² die es erlauben die Annahmen über die funktionalen Formen der Regressionen und der Kontrollfunktion stark zu flexibilisieren.

nicht vom üblichen Fluch der Dimensionalität betroffen sind. Ausserdem diskutiert die Arbeit die Erweiterungen auf den Fall mehrerer diskreter oder kontinuierlicher Instrumente und diskreten Ausprägungen der Variablen S .

²² Siehe z. B. Heckman (1990) und Andrews, Schafgans (1998) und auch die Übersichtsartikel von Vella (1998) und Vella and Verbeek (1999).

Wichtig ist in diesem Zusammenhang die Arbeit von Vytlacil (2002), der zeigt, dass die nichtparametrische Identifikation von Selektionsmodellen und Instrumentalvariablenmodellen exakt die gleichen Annahmen verlangt und somit von beiden Modelltypen die gleichen kausalen Effekte identifiziert werden. Mit anderen Worten, nichtparametrische Selektionsmodelle und Instrumentalvariablenmodelle sind identisch. Insbesondere ist das Pendant zur Ausschlussrestriktion beim IV-Modell die Bedingung, dass die Kontrollfunktion mit einer Variablen variieren muss, die nicht in der Regressionsfunktion enthalten ist.

2.3.4 Im Datensatz fehlen wichtige Variablen II: Zeitkonstante Verzerrung und das regression discontinuity design

Ist der Datensatz nicht informativ genug, um die Endogenität / Selektivität kontrollieren zu können, so haben in der Literatur weitere alternative Annahmen Verwendung gefunden, von denen zwei im folgenden kurz beschrieben werden.

Die erste Annahme geht davon aus, dass man sich in einer Situation befindet, in der man von einer zeitlichen Konstanz der Selektionsverzerrung ausgehen kann, das heisst die potentiellen Ergebnisse der Teilnehmer und Nichtteilnehmer weisen den gleichen Trend auf. Mit Daten für die Zeit vor der Teilnahme lässt sich die Selektionsverzerrung korrigieren, falls die Individuen zu diesem Zeitpunkt ihr Verhalten noch nicht im Hinblick auf eine erwartete Teilnahme verändert haben. Man schätzt mit den üblichen Methoden, z. B. Matching in Fitzenberger, Speckesser (2001), oder Eichler, Lechner (2002), den ‚Effekt‘ der Teilnahme auf eine Ergebnisvariable. Da diese Variable u. o. g. Annahme noch gar nicht von der späteren Teilnahme beeinflusst sein kann, kann das Schätzergebnis als Abschätzung der verbleibenden Selektionsverzerrung interpretiert werden. Nun wiederholt man die Schätzung nach der Teilnahme und korrigiert das Ergebnis um die zuvor geschätzte Verzerrung. In der Literatur sind inzwischen viele unterschiedliche Varianten dieser sogenannten Differenzen-in-Differenzen-Verfahren bekannt, deren detaillierte Diskussion den Rahmen dieser Arbeit sprengen würden.²³

²³ Siehe z.B. Bertrand, Duflo, Mullainathan (2001), Heckman, Hotz (1989), Heckman, Ichimura, Todd (1997), Heckman, Robb (1986), Meyer (1995), Mullahy (1999), Rosenbaum (2001).

Ein ähnlicher Weg Homogenität auszunutzen, jetzt allerdings nicht in der zeitlichen Dimension sondern im Querschnitt, benutzt das sogenannte *regression discontinuity design*, das auf Campbell, Stanley (1963) und Thistlewaite, Campbell (1960) zurückzuführen ist und inzwischen häufig angewandt wurde.²⁴ Zentral für diese Idee ist die Existenz einer kontinuierlichen Variablen, die sich in zwei Wertebereiche aufteilen lässt. Ein Wertebereich ist mit einer Teilnahme, der andere Wertebereich mit einer Nichtteilnahme verbunden. Denkt man an Regressionsfunktionen, dann sollte die Regressionsfunktion am Übergang der beiden Wertebereiche einen Sprung aufweisen, daher der Name *regression discontinuity design*. Allgemeiner geht es darum, dass Personen, die sehr ähnliche Charakteristika wie die Teilnehmer aufweisen, aber nicht teilnehmen durften, eine gute Vergleichsgruppe darstellen könnten. Wird zum Beispiel ein Arbeitsmarktprogramm für Jugendliche bis 24 Jahre aufgelegt, so eignet sich möglicherweise die Gruppe der 25-jährigen gut als Vergleichsgruppe.

2.3.5 Im Datensatz fehlen wichtige Variablen III: Schranken

Was bleibt, wenn man sich eingestehen muss, dass es für die interessierende Fragestellung nicht möglich ist, mit dem zur Verfügung stehendem Datenmaterial kausale Effekte zu schätzen? Manski (1990) und Robins (1989) zeigen, dass selbst in diesem Falle weitere Informationen aus den Daten gewonnen werden können. Zwar ist der kausale Effekt nicht identifiziert, jedoch kann seine Grössenordnung zumindest eingegrenzt werden, falls die Ergebnisvariable einen begrenzten Wertebereich hat. Für das Beispiel einer Indikatorvariablen für Y (z. B. beschäftigt vs. nicht beschäftigt), ergeben sich für ATET aus dessen Definition in Gleichung (3) ohne weiteres Wissen über den Selektionsprozess folgende mögliche Werte:

$$\theta_0 \in [E(Y | S = 1) - 1, E(Y | S = 1)]. \quad (12)$$

Das offensichtliche Problem aus Sicht der wirtschaftspolitischen Schlussfolgerungen besteht darin, dass Null als Effekt der Massnahme nicht ausgeschlossen werden kann. Aus diesem Grund wurden in einigen

²⁴ Interessante Anwendungen sind z. B. Angrist, Lavy (1998, 1999), und Berk, Leeuw (1999). Hahn, Todd, van der Klaaw (2001) leiten rigoros die für die nichtparametrische Identifikation notwendigen Bedingungen her.

Papieren verschiedene Annahmen vorgeschlagen, die zwar alle nicht restriktiv genug sind, die kausalen Effekte zu identifizieren, die aber ausreichen, um die Intervallbreite weiter zu reduzieren.²⁵

3 Daten

Allen ökonometrischen Verfahren ist gemein, dass sie erst dann überzeugend für die wirtschaftspolitische Anwendung sind, wenn sie auf einer informativen Datenbasis beruhen. In der Vergangenheit war sicher neben dem Desinteresse der jeweiligen Regierung die in vielen Ländern schlechte Datenlage einer der Hauptgründe für den fehlenden wirtschaftspolitischen Widerhall mikroökonomischen Evaluationsstudien. Entsprechend besteht ein wesentlicher Bestandteil der aktuellen Evaluationsanstrengungen in Deutschland und anderen europäischen Ländern darin, die Datenbasis dadurch zu verbessern, dass zum einen mehr individuelle Informationen über potentielle Teilnehmer an den einzelnen Massnahmen zur Verfügung gestellt werden und zum andern der Datenzugang für Evaluatoren einfacher wird (insbesondere durch gute Aufbereitung der Daten). Die neuen Datensätze kommen im wesentlichen aus Verknüpfungen einzelner existierender administrativer Individualdatensätze zustande, die aber in einigen Fällen, insbesondere in Grossbritannien, durch damit verknüpfte gezielte Umfragen ergänzt werden. Derartige Umfragen erlauben es, im Datensatz ausgemachte Informationslücken entsprechend zu schliessen. Obwohl schon erhebliche Fortschritte bei der Datenerhebung und Aufbereitung gemacht wurden, sind die entsprechenden Projekte in den einzelnen Ländern noch nicht abgeschlossen.

Nachdem auch in Deutschland Interesse an einer wissenschaftlich zufriedenstellenden Evaluation der AAMP entstand, sind Anstrengungen zu beobachten, bessere Daten zur Verfügung zu stellen. Diese neue Datenbasis wird es hoffentlich erlauben, die zuvor diskutierten Methoden für eine Verbesserung der wirtschaftspolitischen Praxis im Bereich AAMP einzusetzen. Jedoch wird es bis zum Vorliegen erster Studien auf Basis der neuen Daten noch einige Zeit dauern.

²⁵ Siehe z.B. Manski (1990, 1997) und Lechner (1999).

4 Einige Ergebnisse bisheriger Studien

Da es zum einen schon eine Vielzahl an Übersichtsarbeiten zu den existierenden empirischen Evaluationsstudien gibt, und zum anderen für viele Länder deren Ergebnisse auf einer oft (noch) zweifelhaften Datenbasis beruhen, wird in diesem Abschnitt zunächst nur ein Überblick über viele, die empirische Literatur zusammenfassende Übersichtsartikel gegeben, bevor einige Ergebnisse für die Schweiz etwas genauer dargestellt werden. Die Schweiz ist für die Situation in Deutschland interessant, da die Arbeitsmärkte institutionell in vielerlei Hinsicht ähnlich sind und die schweizerischen Evaluationsstudien auf einer recht soliden Datenbasis beruhen.

International

Mikroökonometrische Evaluationen AAMP unterschiedlichster Qualität und Politikrelevanz gibt es inzwischen in Hülle und Fülle. International vergleichend bieten die Arbeiten, die im Rahmen und Auftrag der OECD und ILO erstellt wurden, den besten Überblick (Fay, 1996, ILO, 1998, Martin, Grubb, 2001).²⁶ Leider werden dort, wie in vielen anderen Überblicksartikeln, nur einzelne Ergebnisse (oft unvollständig) zusammengetragen und vergleichend dargestellt. Was vollkommen fehlt, ist die kritische Bewertung der jeweils angewandten Methodik und des zugrundeliegenden Datenmaterials.

Da länderspezifische Zusammenfassungen mehr Platz für die einzelne Studie verwenden, und auch entsprechenden Autoren sich genauer mit dem Land befassen konnten, sind diese Arbeiten oft von grösserem Nutzen. Insbesondere für die USA gibt es sehr gute Übersichtsarbeiten, die sich im wesentlichen auf die Ergebnisse experimenteller Studien beziehen und somit einen besonderen Vertrauensbonus geniessen (siehe z. B. Friedlander, Greenberg, 1997, LaLonde, 1995, US-Department of Labor, 1995). Insgesamt ist der Tenor dieser Studien eher negativ, da kaum systematisch positive Effekte der unterschiedlichen Massnahmen gefunden werden konnten. Bestenfalls gibt es positive Effekte für einzelne Untergruppen. Die Übertragbarkeit dieser Ergebnisse auf die deutschen (und schweizerischen) Verhältnisse ist allerdings

problematisch, da zum einen die Ergebnisse grösstenteils aus den 80'iger Jahren stammen, und zum anderen der US-Arbeitsmarkt hinsichtlich vieler Elemente nur sehr begrenzt mit dem deutschen und schweizerischen Arbeitsmarkt vergleichbar ist. Entsprechend ist auch nicht klar, dass man bei einer positiven oder negativen Wirkung eines Programms in den USA auf eine gleichartige Wirkung in einem andersartigen Arbeitsmarkt schliessen kann.

In Schweden und Grossbritannien empfiehlt es sich wahrscheinlich, mit den Schlussfolgerungen noch etwas zu warten, bis die neu aufgebauten oder im Aufbau begriffenen Datenbasen für hinreichend viele moderne Evaluationsstudien zur Verwendung gekommen sind. Der Stand des Wissens für Grossbritannien wird gut in Dolton, Balfour (2002) wiedergegeben. Neuere Ergebnisse für Schweden finden sich zum Beispiel in Carling, Richardson (2001), Calmfors, Forslund, Helmstrom (2002), und Acknell Agell, Forslund, Helmström, Nordström, Runeson, Öckert (2002).

Deutschland

Die Art, das Ausmass und die vorläufigen, völlig unzureichenden Erkenntnisse über die Effektivität der AAMP, die in jüngster Zeit in Deutschland durchgeführt wurde, ist inzwischen erschöpfend dokumentiert.²⁷ Zusammenfassend lässt sich feststellen, dass sowohl der Umfang der AAMP als auch die Anzahl der erstellten mikroökonomischen Evaluationsstudien beeindruckend ist. Die Lehre aus diesen Studien ist ernüchternd: Zwar verzeichnen einige Studien durchaus interessante methodische Fortschritte, jedoch sind die sich oft widersprechenden wirtschaftspolitischen Schlussfolgerungen weitgehend nutzlos. Grund dafür ist sicher nicht die Inkompetenz der Evaluatoren, sondern vielmehr die bisher schlechte Datenlage in Deutschland. Keiner der Wissenschaft und den Wirtschaftsforschungsinstituten zur Verfügung stehenden Sekundärdatensätze ist ausreichend bezüglich Grösse und Informationsgehalt, um sowohl unterschiedliche

²⁶ Andere Übersichten, die allerdings nicht mehr ganz aktuell sind, finden sich in Björklund, Regnér (1996) und Heckman, LaLonde, Smith (1999). Für Osteuropa siehe beispielsweise die Arbeiten von O'Leary, Kolodziejczyk, Lázár (1998). Die aktuellste Übersicht kann in Steiner und Hagen (2002) gefunden werden.

²⁷ Die Arbeiten von Fitzenberger, Hujer (2002), Fitzenberger, Speckesser (2000), Hujer, Caliendo (2000), Schmidt (2000), und Eichler, Lechner (2001) geben einen vollständigen Überblick über die Massnahmen, die Methoden, die Datengrundlagen und die Ergebnisse für Deutschland.

Massnahmenarten hinreichend fein zu unterscheiden, als auch die Selektion in die Massnahmen angemessen zu berücksichtigen. Sind die Daten nicht informativ genug, kann die beste Ökonometrie nicht weiterhelfen, und die empirischen Ergebnisse sind notgedrungen wirtschaftspolitisch wenig relevant. Inzwischen setzt sich anscheinend zumindest in Teilen der öffentlichen Verwaltung die Einsicht durch, dass man von wirtschaftspolitisch relevanten mikroökonomischen Evaluationsstudien am Ende nur profitieren kann.

Schweiz

Mit der starken Expansion der AAMP ab 1996 wurde gleichzeitig das gesamtschweizerische Evaluationsprogramm beschlossen. Inzwischen liegen sechs Evaluationsstudien vor, die verschiedene Aspekte der Massnahmen beleuchten. Zwei der Studien sind (mikro-) ökonomischer Natur und beschäftigen sich im wesentlichen mit der Abschätzung der Wirkungen der AAMP auf die späteren individuellen Erwerbswahrscheinlichkeiten.²⁸ Im Vergleich zu den Studien für Deutschland besteht ein grosser Vorteil dieser Arbeiten in der recht guten Datenbasis: Administrative Daten der Arbeitslosenversicherung wurden mit Daten der Alters- und Hinterbliebenenversicherung (entspricht der deutschen Rentenversicherung) zu einer grossen Stichprobe auf individueller Basis verbunden. Obwohl auch dieser Datensatz noch Wünsche offen lässt, war es weitgehend möglich, sowohl die heterogenen Programmtypen als auch die Selektion in die Massnahmen zu berücksichtigen.²⁹

Die Studien zeigen, dass die verschiedenen Massnahmen höchst unterschiedlich wirken. Insbesondere finden die Arbeiten von Gerfin, Lechner (2002) und Gerfin, Lechner, Steiger (2002), dass die sogenannten Zwischenverdienste alle anderen Massnahmen dominieren. Ein Zwischenverdienst ist eine Erwerbstätigkeit, die jemand vorübergehend annimmt und deren Lohn weniger als die zustehende Arbeitslosenentschädigung beträgt. Personen im Zwischenverdienst müssen alle Kontrollvorschriften einhalten, Stellenbemühungen unternehmen und vermittelbar sein. Die Arbeitslosenversicherung leistet an Personen im Zwischenverdienst sogenannte "Kompensationszahlungen", welche 70 oder 80 Prozent (je nach der gel-

²⁸ Siehe Gerfin und Lechner (2002), Lalive, van Ours und Zweimüller (2000), Lechner und Gerfin (2000) und die Zusammenfassungen in Curti und Zürcher (2000).

²⁹ Für eine Beschreibung der Daten siehe Gerfin, Lechner (2002) und Lalive, van Ours, Zweimüller (2000).

tenden Ersatzrate des Arbeitslosengelds) der Differenz zwischen dem Zwischenverdienst und dem versicherten Verdienst ausmachen.³⁰ Eine weiterführende Analyse für verschiedene Untergruppen zeigt, dass der Zwischenverdienst bei Personengruppen besonders effektiv wirkt, die lange Arbeitslosigkeit, geringe Qualifikation und schlechte Vermittelbarkeit aufweisen. Zusammenfassend lässt sich festhalten, dass nach diesen Analysen Zwischenverdienste als einziges Programm einen klar positiven Effekt auf die Wiederbeschäftigungschancen aufweisen, während die klassischen Beschäftigungsprogramme ähnlich zu den deutschen ABM zumindest kurzfristig, wahrscheinlich aber auch mittelfristig negativ wirken. Für die übrigen Bestandteile der AAMP der Schweiz lassen sich noch keine so eindeutigen Ergebnisse finden.³¹

5 Schlussfolgerung

Die moderne mikroökonometrische Evaluationsliteratur hat in ihrem methodischen Teil die Frage nach den Bedingungen für eine (nichtparametrische) Identifikation heterogener kausaler Effekte in den Vordergrund gestellt. Damit wurden auch althergebrachten Schätzverfahren zu neuer Bedeutung und zu einer neuen Interpretation verholfen. Sozusagen als Nebenprodukte entstanden Schätzverfahren, die sich unter gewissen Bedingungen hervorragend zur nichtparametrischen Schätzung von Durchschnittseffekten eignen können. Es ist anzunehmen, dass sich das in dieser Literatur entwickelte Denkmuster und auch einige der Schätzmethoden im Mainstream der angewandten (Mikro- und Panel-) Ökonometrie der Zukunft etablieren werden. Immerhin es das Ziel eines jeden angewandten ökonometrischen Papiers kausale Schlussfolgerungen zu ziehen.

Die insbesondere in Grossbritannien, Schweden, der Schweiz und auch in Deutschland entstehenden Datensätze deuten auf eine in naher Zukunft zu erwartende Phase vielfältiger und glaubwürdiger neuer inhaltlicher Erkenntnisse über die Wirksamkeit der einzelnen Komponenten der AAMP auf die jeweils davon betroffenen Arbeitslosen hin.

³⁰ Der versicherte Verdienst definiert das für die Berechnung des Arbeitslosengeldes relevante Einkommen.

³¹ Inzwischen wird das Datenmaterial verbessert, so dass deutlich weitergehende Analysen möglich sein werden.

Abschliessend noch eine Anmerkung zu einem Thema, das bis jetzt in diesem Papier noch nicht berührt wurde: Untersuchungen auf Makroebene. Ihnen kann potentiell eine wichtige Bedeutung zukommen.³² Die neueren empirischen Analysen auf aggregierter Ebene basieren im wesentlichen entweder auf einer Untersuchung der Job-Matching Effizienz auf einem wohl definierten 'Arbeitsmarkt' (siehe z.B. Büttner und Prey, 1997, Puhani, 1999), oder auf ökonometrisch untermauerten berechenbaren Gleichgewichtsmodellen (siehe z. B. Heckman, 1999). Es ist allerdings in diesem Bereich augenscheinlich bislang nicht annähernd ein Konsens über das beste Vorgehen zu erkennen, da alle Ansätze mit vielen konzeptionellen und Umsetzungsproblemen behaftet sind. Es sei auch darauf hingewiesen, dass derartigen Studien eher für die Frage der Gesamtgrösse einzelner Massnahmen relevant sind als zum Beispiel für die Entscheidung, welche Typen von Arbeitslosen an welchen Programmen teilnehmen sollten.

Literaturverzeichnis

- Abadie, A., J. Angrist, und G. Imbens (2000): "Instrumental Variable Estimates of the Effects of Subsidized Training on the Quantiles of Trainee Earnings", *Econometrica*, 70, 91-117.
- Acknell Agell, S., A. Forslund, M. Helmström, O. Nordström, C. Runeson, und B. Öckert (2002): "Follow-up of EU recommendations on labour market policies", IFAU report.
- Andrews, D. W. K., und M. A. Schafgans (1998): "Semiparametric estimation of the intercept of a sample selection model", *The Review of Economic Studies*, 65, 497-517.
- Angrist, J. D., und A. B. Krueger (2001): "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments", *Journal of Economic Perspectives*, 15, 69-85.
- Angrist, J. D., und V. Lavy (1998): "Using Maimonides Rule to Estimate the Effects of Class Size on Scholastic Achievements", *Quarterly Journal of Economics*.
- Angrist, J. D., und V. Lavy (1999): "New Evidence on Classroom Computers and Pupil Learning", NBER working paper 7424.
- Angrist, J.D., G.W. Imbens, und D.B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables", *Journal of the American Statistical Association*, 91, 444-472, with discussion.
- Angrist, J. D., und A. B. Krueger (1999): "Empirical Strategies in Labor Economics", in O. Ashenfelter und D. Card (Hrsg.), *Handbook of Labor Economics*, Vol. III A, Kapitel 23, 1277-1366.

³² Siehe z. B. die Erwägungen in Garfunkel, Manski, Michalopoulos (1992), Dahlberg, Forslund (1999) und Forslund, Kolm (2000).

- Berk, R.A., und J. De Leeuw (1999): "An Evaluation of California's Inmate Classification System Using a Generalised Regression Discontinuity Design", *JASA*, 94, 1045-1052.
- Bertrand, M., E. Duflo, und S. Mullainathan (2001): "How much should we trust differences-in-differences estimates", NBER WP 8841
- Björklund, A. und H. Regnér (1996): "Experimental Evaluation of European Labour Market Policy", in: Schmid, G., J. O'Reilly, und K. Schömann (Hrsg., WZB, 1996): *International Handbook of Labor Market Policy and Evaluation*, Cheltenham: Edward Elgar, Chapter 3, 89-114.
- Burtless, G. (1995): „The Case for Randomized Field Trials in Economic and Policy Research.“ *Journal of Economic Perspectives*, 9(2), 63-84.
- Büttner, T., und H. Prey (1997): "Does Active Labour Market Policy Affect Structural Unemployment? An Empirical Investigation for West German Regions, 1986-1993", *ZWS*, 118, 389-413.
- Calmfors, L., A. Forslund, und M. Helmstrom (2002): "Does active labour market policy work? Lessons from the Swedish experience", IFAU working paper 2002-4.
- Campbell, D. und J. Stanley (1963): „Experimental and quasi-experimental designs for research on teaching“, in: N. Gage, Hrsg., *Handbook of research on teaching* (Rand McNally, Chicago), 171-246.
- Carling, K., und K. Richardson (2001): "The relative efficiency of labor market programs: Swedish experience from the 1990's", *IFAU discussion papers*, 2001:2.
- Chernozhukov, V., und Ch. Hansen (2001): "An IV Model of Quantile Treatment Effects", *mimeo*.
- Cox, D.R (1992): "Causality: Some Statistical Aspects", *Journal of the Royal Statistical Society, A*, 155, 291-301.
- Curti, M., und B. Zürcher (2001): "Die gesamtschweizerische Evaluation der aktiven Arbeitsmarktpolitik", *Die Volkswirtschaft*, 4/2000, 6-10.
- Dahlberg, M., und A. Forslund (1999): "Direct displacement effects of labour market programmes: the case of Sweden", *IFAU DP*, 1999/7.
- Dawid, A. P. (2000): "Causal Inference Without Counterfactuals", *Journal of the American Statistical Society*, 95, 407-448. With discussion.
- Dehejia, R. H. und S. Wahba (1999): "Causal Effects in Non-experimental Studies: Reevaluating the Evaluation of Training Programmes", *Journal of the American Statistical Association*, 94, 1053-1062.
- Dehejia, R. H. und S. Wahba (2002): "Propensity score-matching methods for nonexperimental causal studies", *Review of Economics and Statistics*, 84, 151-161.
- Dolton, P., und Y. Balfour (2002): "Der New Deal, "Welfare to Work"-Programme in Grossbritannien“, *Perspektiven der Wirtschaftspolitik*, 3, 175-187.
- Durbin, J. (1954): "Errors in Variables", *Review of the International Institute*, 22, 23-32.
- Eichler, M., und M. Lechner (2001): "Die aktive Arbeitsmarktpolitik in den neuen Bundesländern", in: J. Rangnitz (Hrsg.), *Zehn Jahre Deutsche Einheit: Bilanz und Perspektiven*, Halle: IWH, 148- 168.

- Eichler, M., und M. Lechner (2002): „An Evaluation of Public Employment Programmes in the East German State of Sachsen-Anhalt“, *Labour Economics: An International Journal*, 9, 143-186.
- Fay, R. G. (1996): "Enhancing the Effectiveness of Active Labour Market Policies: Evidence from Programme Evaluations in OECD countries", *Labour Market and Social Policy Occasional Papers*, #18, OECD.
- Fechner, G.T. (1860): *Elemente der Psychophysik*, Leipzig: Breitkopf und Härtel.
- Fitzenberger, B., und S. Speckesser (2000): "Zur wissenschaftlichen Evaluation der Aktiven Arbeitsmarktpolitik in Deutschland: Ein Überblick", erscheinen in: Bellmann, L., und R. Hujer (Hrsg.), *Beiträge zur Arbeitsmarkt- und Berufsforschung* (BeitrAB), 33, IAB, Nürnberg.
- Fitzenberger, B., und S. Speckesser (2001): "Comparing Difference-in-Difference Estimators based on Nearest Neighbor or Kernel Matching: An Evaluation of Training in East Germany," mimeo.
- Forslund, A., und A. Kolm (2000): "Active labour market policies and real-wage determination – Swedish evidence", IFAU WP 2000:7.
- Friedlander, D., und D. H. Greenberg (1997): "Evaluating Government Training Programs for the Economically Disadvantaged", *Journal of Economic Literature*, 35, 1809-1855.
- Frölich, M. (2002a), *Programme Evaluation and Treatment Choice*, Dissertation an der Universität St. Gallen, im erscheinen, Heidelberg: Springer Verlag.
- Frölich, M. (2002b): „Nonparametric IV estimation of local average treatment effects with covariates“, mimeo.
- Garfinkel, I., C.F. Manski und C. Michalopoulos (1992): "Micro Experiments and Macro Effects", in: C.F. Manski und I. Garfinkel (Hrsg.), *Evaluating Welfare and Training Programs*, Cambridge: Harvard University Press, 253-273.
- Gerfin, M., und M. Lechner (2002): "Microeconomic Evaluation of the Active Labour Market Policy in Switzerland," *The Economic Journal*, im Erscheinen.
- Greenberg, D., und M. Shroder (1997): *The Digest of Social Experiments*, 2nd edition, The Urban Insitute Press.
- Hahn, J., P. Todd, und W. van der Klaaw (2001): "Identification and Estimation of Treatment Effects Using a Regression Discontinuity Design", *Econometrica*, 69, 201-209.
- Heckman, J. J. (1979): "Sample Selection as a Specification Error", 47, *Econometrica*, 153-161.
- Heckman, J. J. (1990): "Varities of Selection Bias", *American Economic Review*, Papers and Proceedings, 313-318.
- Heckman, J. J. (1999): "Accounting for heterogeneity, diversity and general equilibrium in evaluating social programs", *NBER working paper* 7230.
- Heckman, J. J., und V. J. Hotz (1989): "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training", *Journal of the American Statistical Association*, 84, 862-.
- Heckman, J. J., und R. Robb (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes", in Howard Wainer (Hrsg.), *Drawing Inferences from Self-Selected Samples*, 63-113.

- Heckman, J. J., und J. A. Smith (1995). „Assessing the Case for Social Experiments.“ *Journal of Economic Perspectives*, 9(2), 85-100.
- Heckman, J. J., und J. A. Smith (1999): "The Pre-programme earnings dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies", *The Economic Journal*, 109.
- Heckman, J. J., und E. Vytlačil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects", *Proceedings of the National Academy of Science*, April 13, 1999.
- Heckman, J. J., und E. Vytlačil (2002): "Econometric Evaluation of Social Programs", erscheint in J. J. Heckman und E. Leamer (Hrsg.), *Handbook of Econometrics*, Volume 6, North-Holland: Amsterdam.
- Heckman, J. J., H. Ichimura, und P. E. Todd (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program", *The Review of Economic Studies*, 64, 605-654.
- Heckman, J. J., R. J. LaLonde und J. A. Smith (1999): "The Economics and Econometrics of Active Labour Market Programs", in O. Ashenfelter und D. Card (Hrsg.), *Handbook of Labor Economics*, Vol. III A, chapter 31, 1865-2097.
- Heckman, J. J., J. A. Smith und N. Clements (1997): "Making the most out of Programme Evaluations of Social Experiments: Accounting for Heterogeneity in Programmes Impacts", *The Review of Economic Studies*, 64, 487-535.
- Heckman, J. J., H. Ichimura, J. Smith und P. Todd (1998): "Characterisation Selection Bias Using Experimental Data", *Econometrica*, 66, 1017-1098.
- Holland, P. W. (1986): "Statistics and Causal Inference", *Journal of the American Statistical Association*, 81, 945-970, mit Diskussion.
- Holland, P. W. (1989): "Comment: It's Very Clear", *Journal of the American Statistical Association*, 84, 875.
- Hujer, R., K. Maurer und M. Wellner (1997): "The Impact of Training on Unemployment Duration in West Germany – Combining a Discrete Hazard-Rate Model with Matching Techniques", *Discussion papers in Economics # 74*, University of Frankfurt.
- Hujer, R., und M. Caliendo (2000): "Evaluation of Active Labour Market Policy", IZA DP 236.
- Hujer, R., und M. Wellner (2000): "The Effects of Public Sector Sponsored Training on Individual Employment Performance in East Germany", *mimeo*.
- ILO (1998): *World Employment Report 1998-1999 – Employability in the Global Economy: How Training Matters*.
- Imbens, G. W. (2000): "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, 87, 706-710.
- Imbens, G. W., und J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, 62, 446-475.
- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (2001): "Wege zu einer besseren informationellen Infrastruktur", *Gutachten im Auftrag des Bundesministeriums für Bildung und Forschung*.

- Lalive, R., J. C. van Ours und J. Zweimüller (2000): "The Impact of Active Labor Market Policies and Benefit Entitlement Rules on the Duration of Unemployment", *mimeo*.
- LaLonde, R. J. (1995): "The Promise of Public Sector-Sponsored Training Programs", *Journal of Economic Perspectives*, 9, 149-168.
- Leamer, E. E., J.W. Pratt, und R. Schlaifer (1988): "On the Interpretation and Observation of Laws", *Journal of Econometrics*, 39, 23-52.
- Leamer, E. E. (1983): "Let's Take the Con out of Econometrics", *American Economic Review*, 73, 31-43.
- Lechner, M. (1998), *Training the East German Labour Force. Microeconomic Evaluations of Continuous Vocational Training after Unification*, Heidelberg: Physica.
- Lechner, M. (1999a): „Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification,“ *Journal of Business & Economic Statistics*, 17, 74-90.
- Lechner, M. (1999b): „Nonparametric Bounds on Employment and Income Effects of Continuous Vocational Training in East Germany“ *The Econometric Journal*, 2, 1-28.
- Lechner, M. (2000): „An Evaluation of Public Sector Sponsored Continuous Vocational Training Programs in East Germany,“ *The Journal of Human Resources*, 35, 347-375.
- Lechner, M. (2001a): "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption", in M. Lechner und F. Pfeiffer (Hrsg.), *Econometric Evaluation of Active Labour Market Policies*, Heidelberg: Physica, 43-58.
- Lechner, M. (2001b): „The Empirical Analysis of East German Fertility after Unification: An Update,“ *European Journal of Population*, 17, 61-74.
- Lechner, M. (2002a): „Eine wirkungsorientierte aktive Arbeitsmarktpolitik in Deutschland und der Schweiz,“ *Perspektiven der Wirtschaftspolitik*, 3, 159-174, 2002.
- Lechner, M. (2002b): "Programme Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labour Market Policies", *Review of Economics and Statistics*, 84, 205-220.
- Lechner, M. (2002c): "Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods", *Journal of the Royal Statistical Society - Series A*, 165, 59-82.
- Lechner, M. (2002d): "Dynamic matching estimation", draft.
- Lechner, M., und F. Pfeiffer (2001, Hrsg.), *Econometric Evaluation of Labour Market Policies*, Heidelberg: Physica.
- Manski, C. F. (1997): "Monotone Treatment Response", *Econometrica*, 65, 1311-1334.
- Manski, C.F. (1990): "Nonparametric Bounds on Treatment Effects", *American Economic Review, Papers and Proceedings*, 80, 319-323.
- Martin, J. P., und D. Grubb (2001): "What works and for whom: a review of OECD countries' experiences with active labour market policies", IFAU WP 2001:14.
- Meyer, B.M. (1995): "Natural and Quasi-Experiments in Economics", *Journal of Business & Economic Statistics*, 13, 151-161.

- Miquel, R., und M. Lechner (2001): "Identification of Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions", mimeo.
- Miquel, R. (2002a): "Identifikation of Effects of Dynamic Treatments by Instrumental Variables", mimeo.
- Miquel, R. (2002b): "Identifikation of Effects of Dynamic Treatments with a Difference-in-Difference Approach", mimeo.
- Mullahy, J. (1999): "Interaction effects and difference-in-difference estimation in loglinear models", *NBER Technical Working Paper* 245.
- Neyman, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9", translated in *Statistical Science* (with discussion), 1990, 5, 465-480.
- Neyman, J. (1935): "Statistical Problems in Agricultural Experimentation", Supplement to *the Journal of the Royal Statistical Society*, 2, 107-180.
- O'Leary, C. J., P. Kolodziejczyk, und G. Lázár (1998): "The net impact of active labour market programmes in Hungary and Poland", *The International Labour Review*, 137, 321-346.
- Pearl, J. (2000), *Causality - Models, Reasoning, and Inference*, Cambridge: Cambridge University Press.
- Persson, T. (2001): "Currency Unions and Trade: How Large is the Treatment Effect", mimeo.
- Persson, T., und G. Tabellini (2001): "Do Constitutions Cause Large Governments? Quasi-experimental evidence", mimeo.
- Robins, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393-1512, with 1987 Errata to "A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect." *Computers and Mathematics with Applications*, 14:917-921; 1987 Addendum to "A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect." *Computers and Mathematics with Applications*, 14:923-945; and 1987 Errata to "Addendum to 'A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect'." *Computers and Mathematics with Applications*, 18:477.
- Robins, J. M. (1989): "The Analysis of Randomized and Nonrandomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies", Sechrest, L., H. Freeman, und A. Mulley (Hrsg.), *Health Service Research Methodology: A Focus on Aids*, 113-159, Washington, D.C.: Public Health Service, National Center for Health Services Research.
- Rosenbaum, P. R. (1984): "The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment", *The Journal of the Royal Statistical Society, Series A*, 147, 656-666.
- Rosenbaum, P. R. (1995), *Observational Studies*, New York: Springer.
- Rosenbaum, P. R. (2001): "Stability in the Absence of Treatment", *Journal of the American Statistical Association*, 96, 210-219.

- Rosenbaum, P.R. und D.B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70, 41-50.
- Rosenbaum, P.R. und D.B. Rubin (1985): "Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score", *The American Statistician*, 39, 33-38.
- Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D. B. (1979): "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies", *Journal of the American Statistical Association*, 74, 318-328.
- Rudolph, H., und M. Müntnich (2001): "Profiling zur Vermeidung von Langzeitarbeitslosigkeit – erste Ergebnisse aus einem Modellprojekt", *MittAB*, 530-553.
- Schmidt, C. M. (2000): "Arbeitsmarktpolitische Massnahmen und ihre Evaluierung: Eine Bestandsaufnahme", *Vierteljahreshefte zur Wirtschaftsforschung*, 69, 425-437.
- Smith, J., und P. Todd (2002): "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators", forthcoming in *Journal of Econometrics*.
- Sobel, M.E. (1994): "Causal Inference in the Social and Behavioral Sciences", in: G. Arminger, C.C. Clogg and M.E. Sobel (Hrsg.): *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, New York: Plenum Press, 1-38.
- Steiner, V., und T. Hagen (2002): "Was kann die Aktive Arbeitsmarktpolitik in Deutschland aus der Evaluationsforschung in anderen europäischen Ländern lernen", *Perspektiven der Wirtschaftspolitik*, 3, 189-206.
- Thistlewaite, D., und D. Campbell (1960): "Regression-discontinuity Analysis: An alternative to the ex post facto experiment", *Journal of Educational Psychology*, 51, 309-317.
- Trochim, W. (1984), *Research Design for Program Evaluation: the Regression Discontinuity Approach*, Beverly Hills: Sage Publications.
- US-Department of Labor, Office of the Chief Economist (1995): What's Working (and what's not) - A Summary of Research on the Economic Impacts of Employment and Training Programs.
- Vella, F. (1998): "Estimating Models with Sample Selection Bias: A Survey", *Journal of Human Resources*, 33, 127-169.
- Vella, F., und M. Verbeek (1999): "Estimating and Interpreting Models with Endogenous Treatment Effects", *Journal of Business & Economic Statistics*, 17, 473-478.
- Vytlacil, E. (2002): „Independence, monotonicity and latent variable models: an equivalence result“, *Econometrica*, 70, 331-341.