

Mikroevaluationen: Bildung von Vergleichsgruppen zur Schätzung individueller Förderwirkungen

Ursula Jaenichen

1 Einleitung

Das Ziel mikroökonomischer Evaluationen arbeitsmarktpolitischer Eingriffe besteht darin, die Wirkungen dieser Eingriffe auf die davon betroffenen Personen oder Personengruppen zu untersuchen. Gefragt wird z. B. nach dem Einfluss eines Programms auf die individuellen Einkommen der geförderten Personen, auf die Wahrscheinlichkeit und die Stabilität einer anschließenden Beschäftigung, auf die Überlebenswahrscheinlichkeit oder das Wachstum von Betrieben.

Um zu beurteilen, wie sich die Förderung durch ein arbeitsmarktpolitisches Instrument auf die weiteren Erwerbschancen der geförderten Personen ausgewirkt hat, wird mit dem Vergleichsgruppenkonzept eine Gruppe nicht geförderter Personen als Referenzmaßstab gewählt. Der Grundgedanke dabei ist, mit der Vergleichsgruppe die Ergebnisse zu simulieren, die die geförderten Personen ohne Förderung erzielt hätten. Während durch die Heranziehung von Vergleichsgruppen prinzipiell eine intuitiv leicht nachvollziehbare Schätzung der individuellen Förderwirkungen ermöglicht wird, kann die Auswahl oder Bildung einer solchen Gruppe „vergleichbarer Personen“ je nach Datenlage und untersuchter Fragestellung mit Schwierigkeiten behaftet sein.

Im Folgenden sollen zunächst Probleme, die bei der Konstruktion von Vergleichsgruppen auftreten, erläutert werden. Anschließend erfolgt eine Darstellung der Grundzüge des „Matching“-Verfahrens. Bedingungen und Kriterien für die Auswahl einer geeigneten Vergleichsgruppe sind expliziter Bestandteil dieser Evaluationstechnik. Abschließend wird ein laufendes IAB-Projekt zur Evaluation betrieblicher Einstellungshilfen vorgestellt.

2 Grundprobleme der Vergleichsgruppenbildung

Einmal angenommen, die Wirkung eines Förderinstruments auf eine bestimmte Ergebnisvariable, z. B. die Wahrscheinlichkeit der Aufnahme einer Beschäftigung für zuvor arbeitslose Personen, soll untersucht werden. Dann greift der zunächst vielleicht nahe liegende Weg, den Erfolg der Förderung durch den Anteil der Personen zu messen, die in einem bestimmten Zeitraum nach Einsetzen der Förderung in Beschäftigung übergegangen sind, zu kurz. Damit würde übersehen, dass ein gewisser Anteil der geförderten Personen

auch ohne Förderung in Beschäftigung gewechselt wäre. Um den Fördererfolg abzuschätzen, muss daher auch eine Antwort auf die Frage gefunden werden, inwieweit die geförderten Personen auch ohne Förderung eine Beschäftigung aufgenommen hätten. Diese Fragestellung ist immer hypothetisch, da für ein und dieselbe Person niemals gleichzeitig die Zustände „gefördert“ und „nicht gefördert“ beobachtet werden können. Die so genannte kontrafaktische Situation, hier die Beschäftigungswahrscheinlichkeit geförderter Personen im Fall der Nichtförderung, muss daher durch Rückgriff auf plausible Annahmen konstruiert werden. Dies ist für empirische Studien nichts Ungewöhnliches, so ist etwa die saisonbereinigte Arbeitslosenquote ein geschätztes Maß dafür, wie sich die Arbeitslosenquote entwickelt hätte, würde die Beschäftigung nicht saisonalen Schwankungen unterliegen. Auf der Personenebene liegt es nahe, auf die Beschäftigungswahrscheinlichkeit von nicht durch die Förderung erfassten Personen zurückzugreifen, um die hypothetische Beschäftigungswahrscheinlichkeit geförderter Personen im Fall der Nichtförderung abzuschätzen. Im Zusammenhang mit Evaluationen wird von einer Vergleichs- oder Kontrollgruppe gesprochen, deren „Ergebnisse“ mit denen der geförderten Gruppe von Teilnehmer/innen kontrastiert werden, um den Fördererfolg abzuschätzen.¹ Die im Folgenden diskutierten methodischen Schwierigkeiten drehen sich um die Frage, unter welchen Bedingungen die Bildung von Vergleichs- oder Kontrollgruppen und die darauf basierende Schätzung des Erfolgs eines Förderinstruments unverzerrte Ergebnisse erwarten lassen.

2.1 Selection Bias

Als zentrales Problem bei der mikroökonomischen Evaluation arbeitsmarktpolitischer Instrumente gilt der *selection bias*, auch Selektionsverzerrung oder Verzerrung durch Stichprobenselektion. Hiermit wird bezeichnet, dass sich geförderte und nicht geförderte Personen häufig durch Merkmale unterscheiden, die Einfluss sowohl auf den Förderzugang bzw. die Teilnahme an einer Maßnahme als auch auf das Ergebnis der Förderung, den Fördererfolg, ausüben. Dies können zum einen sehr einfach zu beobachtende Merkmale sein. Zur Illustration sei unterstellt, dass die potenziellen Teilnehmer/innen an einer Maßnahme sich nur durch das Merkmal Geschlecht unterscheiden, Frauen aber im Vergleich zu Männern eine geringere Wahrscheinlichkeit der Förderung aufweisen und gleichzeitig von der Förderung in geringerem Maße profitieren als Männer. Wird die Wirkung der Förderung durch einen einfachen Vergleich der Ergebnisse von Geförderten

¹ Der Begriff der Kontrollgruppe wird hier für die Personengruppe verwendet, die im Rahmen eines Experimentes durch Zufallsauswahl von der Förderung durch ein arbeitsmarktpolitisches Instrument ausgeschlossen wird. Mit Vergleichsgruppe wird eine erst nach Entscheidung über den Förderzugang ausgewählte Gruppe nicht geförderter Personen bezeichnet, ohne dass der Zugang in die Förderung durch ein Experiment gesteuert wurde.

und nicht geförderten Personen geschätzt, ohne dass die geschlechtsspezifische Struktur beider Gruppen Berücksichtigung findet, kommt es zu einer Überschätzung des Fördererfolgs. In diesem Beispiel könnte zur Abschätzung des durchschnittlichen Fördererfolgs die Vergleichsgruppe so gebildet werden, dass der Frauenanteil mit dem der Gruppe der geförderten Personen übereinstimmt. Alternativ könnten getrennte Schätzungen des Fördererfolgs für Männer und Frauen durchgeführt werden. Im allgemeineren Fall, in dem mehrere Merkmale sowohl den Zugang in die Förderung als auch den Fördererfolg beeinflussen, müssen Schätzverfahren gefunden werden, die Unterschiede in den Förderwahrscheinlichkeiten von geförderten und nicht geförderten Personen berücksichtigen.

Erschwert wird die Behandlung des *selection bias* dann, wenn es relevante Merkmale gibt, die nur schwer oder gar nicht zu beobachten sind oder bedingt durch die Datenlage nicht zur Verfügung stehen. Faktoren, die solche unbeobachtete Heterogenität begründen können, sind z. B. Motivation oder durch den Grad schulischer und beruflicher Ausbildung nicht hinreichend abgebildete individuelle Fähigkeiten. Wird auf nicht speziell für Evaluationszwecke erhobene Daten zurückgegriffen, sind häufig wichtige, auf den Förderzugang Einfluss nehmende Merkmale nicht verfügbar. Hierfür kann als Beispiel ein Förderinstrument gelten, das vorrangig in Regionen mit angespannter Arbeitsmarktsituation eingesetzt wird. Wird die Vergleichsgruppe aus einer repräsentativen Umfrage ohne hinreichend differenzierte Regionalinformation gewonnen, wird Selektionsverzerrung aufgrund der unterschiedlichen regionalen Zusammensetzung der Gruppen von geförderten und nicht geförderten Personen resultieren.

2.2 Experimente

Vor allem in den USA wurden bereits häufiger arbeitsmarktpolitische Instrumente in Form von Experimenten umgesetzt, wobei sich hieraus eine einfache und zuverlässige Abschätzung der Wirkungen dieser Instrumente versprochen wurde. Auch die jüngere mikroökonomische Evaluationsliteratur entwickelte sich zum Teil in enger Auseinandersetzung mit den Ergebnissen solcher Experimente, indem nichtexperimentelle Verfahren durch die Anwendung auf experimentelle Daten getestet wurden.

Bei Experimenten soll durch die zufällige Auswahl von Teilnehmer/innengruppe und Kontrollgruppe aus derselben Population garantiert werden, dass kein *selection bias* auftritt. Im Idealfall lässt sich dann der durchschnittliche Fördererfolg durch Vergleich der Mittelwerte der Ergebnisse von Teilnehmer/innen und Kontrollgruppe bestimmen.

Probleme bereitet hierbei jedoch erstens, dass als Teilnehmer/innen ausgewählte Personen die Teilnahme am Experiment verweigern können (vgl. zum Folgenden z. B. Heckman/LaLonde/Smith 1999). Systematische Unterschiede zwischen nicht teilnehmenden und teilnehmenden Personen der ausgewählten Teilnehmer/innengruppe können dann

wieder eine Selektionsverzerrung des geschätzten Fördererfolgs bewirken. Ein möglicher Ausweg besteht darin, nicht die Wirkung des Förderinstruments, sondern die Wirkung des „Angebots der Förderung“ zum Gegenstand der experimentellen Evaluation zu machen.

Ein zweites wichtiges Problem, der *substitution bias*, resultiert daraus, dass die Kontrollgruppe bereits vor Beginn des Experiments festgelegt wird. Nehmen Personen der Kontrollgruppe während des Evaluationszeitraums alternative Fördermöglichkeiten wahr, dann können die Ergebnisse der Kontrollgruppe nicht mehr zur Konstruktion der kontrafaktischen Situation „welche Ergebnisse hätten die Teilnehmer/innen im Fall der Nichtteilnahme aufgewiesen“ verwendet werden. Handelt es sich um ein eher kleines Experiment ohne bedeutsamen Einfluss auf das bestehende Förderspektrum, können die Kontrollgruppenergebnisse in einem solchen Fall zur Beantwortung der Frage „welche Ergebnisse hätten die Teilnehmer/innen gehabt, hätte es nur die sonstigen Fördermöglichkeiten gegeben“ beitragen. Unter Umständen kann hier auch die Evaluationsfragestellung auf das Timing der Förderung ausgerichtet werden, indem die Kontrollgruppenergebnisse als Antwort auf die Frage „welche Ergebnisse hätten die Teilnehmer/innen gehabt, wären sie *noch nicht* gefördert worden“ verwendet werden.²

Durch Experimente mit Zufallsauswahl von Teilnehmer/innen und Kontrollgruppe wird häufig die Zusammensetzung des durch ein Instrument geförderten Personenkreises im Vergleich zum nichtexperimentellen Fall modifiziert. So wird für die USA angeführt, dass einzelne Regionen die Teilnahme an Experimenten verweigerten, weil sie eine Verschlechterung der „Qualität“ unter den geförderten Personen befürchteten. Auch können bei großen Experimenten zur Erzielung der nötigen Fallzahlen zusätzliche Rekrutierungen erforderlich sein, wodurch unter Umständen die Kriterien für den Förderzugang verändert werden. In diesen Fällen von *randomization bias* lassen sich die Ergebnisse von Experimenten unter Umständen nicht verallgemeinern. Der Vorteil der Zufallsauswahl beim Experiment besteht darin, dass die Existenz einer Vergleichsgruppe gesichert wird. Geht es z. B. um ein Förderinstrument, für das eine bestimmte Personengruppe eine Zugangswahrscheinlichkeit von Eins aufweist, wäre ohne den durch die experimentelle Auswahl bewirkten Ausschluss eines Teils dieser Personen das Auffinden einer Vergleichsgruppe nicht möglich.

Zusammengenommen verbessern Experimente bei entsprechendem Design die Datengrundlage für Evaluationen, indem gesichert wird, dass potenzielle Teilnehmer/innen und

² Für eine - allerdings nichtexperimentelle - Evaluation von Instrumenten aktiver Arbeitsmarktpolitik in Schweden mit derartiger Fragestellung siehe Sianesi (2001).

Nichtteilnehmer/innen denselben Personengruppen entstammen. Ferner wird durch die Vorauswahl beider Gruppen die Bildung einer personenbezogenen Datenbasis mit vergleichbaren Informationen erleichtert, Informationen für die Vergleichsgruppe müssen nicht mühsam aus externen Datenquellen konstruiert werden. Da jedoch weder die Teilnahme der potenziellen Fördergruppe noch - sofern es alternative Fördermöglichkeiten gibt - die Nichtteilnahme der Kontrollgruppe erzwungen werden kann, ist auch Evaluation mit experimentellen Daten keine triviale Aufgabe.

2.3 Dynamische Aspekte

Die Zufallsauswahl in Experimenten teilt eine Population zu einem bestimmten Zeitpunkt in eine Teilnehmer/innengruppe und eine Kontrollgruppe ein. Bei der Konstruktion einer Vergleichsgruppe aus nichtexperimentellen Daten wird üblicherweise auf Vergleichbarkeit vor Beginn der Förderung abgestellt. Damit ist jedoch nicht zwingend Vergleichbarkeit der Ergebnisse zu jedem beliebigen Zeitpunkt garantiert.

Ein Problem, das in der Literatur als „*Ashenfelter's Dip*“ (nach der Studie von Ashenfelter 1978) bezeichnet wird, stellt sich, wenn das Erwerbsverhalten der geförderten Personen nicht erst ab Beginn der Förderung, sondern bereits vorher durch die eventuelle Teilnahme beeinflusst wird. Beispielsweise können Personen in Erwartung der Teilnahme an einer Fördermaßnahme ihre Arbeitssuche verringern. Wenn Voraussetzung für eine oder mehrere Förderarten eine bestimmte Mindestarbeitslosigkeitsdauer ist, kann dies bewirken, dass längere Arbeitslosigkeitszeiten auftreten, als es ohne die Aussicht auf Förderung der Fall wäre. Bei der Vergleichsgruppenbildung werden dann unter Umständen die relativen Beschäftigungschancen der geförderten Personen unterschätzt.

Wenn geförderte Personen während der Teilnahme an einem Programm die Intensität der Beschäftigungssuche reduzieren, stellen sich die Fördererfolge, gemessen etwa an der Beschäftigungswahrscheinlichkeit, eventuell erst nach Förderung ein. Bei zu kurzem Beobachtungsfenster kann es dann zu einer Unterschätzung des Fördererfolgs kommen. Wird andererseits die Evaluation auf den Zeitraum nach der Teilnahme beschränkt, bleibt die durch die Teilnahme bewirkte Verringerung der Suchintensität als möglicher negativer Bestandteil des Förderergebnisses unberücksichtigt (siehe hierzu auch die Analysen von Lechner/Gerfin 2000). Für die Gruppe der Vergleichspersonen kann, da hier kein Maßnahme- oder Förderbeginn vorliegt, die Frage, zu oder ab welchem Zeitpunkt die Ergebnisse gemessen werden sollen, ein Problem darstellen.

Werden Ergebnisvariablen in Form von Dauern oder Übergangsraten untersucht, wird sich in den beobachteten Ergebnissen neben der Wirkung der Förderung auch die Wirkung individueller Heterogenität äußern. Auch wenn sich die geförderten Personen und Vergleichspersonen vor Einsetzen der Förderung hinsichtlich ihrer Zusammensetzung

gleichen, wird eine positive Wirkung der Teilnahme dazu führen, dass sich die Merkmalsstrukturen von geförderten Personen und Vergleichspersonen, die im Zeitablauf z. B. als arbeitslos beobachtet werden, auseinanderentwickeln. Weiterhin sind mögliche Dauerabhängigkeiten in den Übergangsraten in Betracht zu ziehen (vgl. Ham/LaLonde 1990). Im Fall von aneinander anschließenden Maßnahmen wie z. B. Weiterbildungsmodulen kann es schwierig sein, den Fördererfolg der nachfolgenden Maßnahmen über den Weg der Vergleichsgruppenbildung zu schätzen.

3 Matching zur Bildung von Vergleichsgruppen

Selektionsverzerrungen von Evaluationsergebnissen entstehen dadurch, dass sich geförderte Personen und potenzielle Vergleichspersonen hinsichtlich ihrer Wahrscheinlichkeit des Förderzugangs unterscheiden und gleichzeitig Unterschiede in der Zugangswahrscheinlichkeit mit Unterschieden in den Förderergebnissen korrelieren. In *matching*-Techniken sind diese unterschiedlichen Zugangswahrscheinlichkeiten das Kriterium für die Vergleichsgruppenbildung.

Handelt es sich nur um wenige Merkmale, von denen Einfluss sowohl auf die Wahrscheinlichkeit des Förderzugangs als auch des Fördererfolgs angenommen werden kann, besteht die einfachste *matching*-Strategie darin, jeder geförderten Person einen statistischen Zwilling, das heißt, eine Person mit identischen Merkmalsausprägungen aus der Gruppe nicht geförderter Personen, zuzuweisen. Ziel ist dabei, eine der Kontrollgruppe im idealtypischen Experiment entsprechende Vergleichsstichprobe zu konstruieren. Das durchschnittliche Förderergebnis kann als einfache Differenz der Durchschnittswerte der Ergebnisvariablen von geförderten Personen und hinzugespielten Vergleichspersonen ermittelt werden. Ferner sind auch Aussagen über den Fördererfolg nach bestimmten Personenmerkmalen oder über die Verteilung der Förderergebnisse ohne großen Aufwand möglich.

Bei einer Vielzahl von zu kontrollierenden Merkmalen kann es jedoch schwierig sein, exakte Zwillinge aufzufinden. Einen Ausweg bietet das *propensity score matching*, das auf ein Ergebnis von Rosenbaum/Rubin (1983) aufbaut. Um die Anzahl der Dimensionen bei der Vergleichsgruppenbildung zu reduzieren, werden zunächst für geförderte und nicht geförderte Personen die Wahrscheinlichkeiten des Förderzugangs in Abhängigkeit von den als relevant erachteten Variablen geschätzt. Die geschätzten Förderwahrscheinlichkeiten bilden die so genannten *propensity scores*. Dann wird das Evaluationsproblem auf den Überlappungsbereich eingegrenzt. Dies ist das Intervall, in dem Werte der *propensity scores* sowohl für geförderte Personen als auch für potenzielle Vergleichspersonen enthalten sind, und sich daher die Verteilungen der *propensity scores* beider Gruppen überlappen. *Matching* kann nun so erfolgen, dass innerhalb dieses Bereichs jeder geför-

der Person eine nicht geförderte Person mit möglichst geringer Differenz in der Förderwahrscheinlichkeit zugewiesen wird.³ Auf diese Art werden die Verteilungen der Förderwahrscheinlichkeiten der geförderten und der nicht geförderten Personen ausbalanciert und auch hier entsprechen unter bestimmten Bedingungen die Eigenschaften der Vergleichsgruppe denen einer experimentellen Kontrollgruppe.

Eine Verallgemeinerung des *matching*-Verfahrens, bei der die paarweise Zuordnung von geförderten Personen und Vergleichspersonen aufgegeben wird, stellt das *kernel matching* dar. Hierbei werden innerhalb des gesamten oder für einzelne Intervalle des Überlappungsbereichs alle potenziellen Vergleichspersonen zur Evaluation der Ergebnisvariablen verwendet, die Distanz zwischen geförderten und nicht geförderten Personen wird durch eine geeignete Gewichtung der Ergebnisvariablen der Vergleichspersonen berücksichtigt (vgl. Heckman u. a. 1999).

Eine zentrale Bedingung, unter der die *matching*-Methode eine unverzerrte Schätzung des durchschnittlichen Förderergebnisses ermöglicht, ist, dass die Kontrolle für die beobachteten, den Zugang in die Förderung beeinflussenden Variablen statistische Unabhängigkeit von Ergebnisvariablen und Förderwahrscheinlichkeiten bewirkt.⁴ Gibt es darüber hinaus unbeobachtete Merkmale, die sowohl den Förderzugang als auch das Förderergebnis beeinflussen, bleibt das Problem der Selektionsverzerrung bestehen.

Die Auswahl der Variablen zur Schätzung der Förderwahrscheinlichkeit ist in praktischen Anwendungen unter Umständen schwierig: So kann eine sehr genaue Vorhersage der Zugangswahrscheinlichkeiten den Überlappungsbereich reduzieren: „... matching requires X variables that are good enough to obtain conditional independence between Y_0 and D, but that are not „too good“ i. e. that predict D perfectly“ (Heckman u. a. 1997: 637). Das mögliche Problem der fehlenden oder nicht vollständigen Überlappung der re-

³ In der Literatur finden sich unterschiedliche *matching*-Algorithmen. Monotone Transformationen der Zugangswahrscheinlichkeiten können ebenfalls (oder besser) als *propensity scores* dienen, alternativ oder in Ergänzung finden statistische Distanzmaße Verwendung. Als *matching*-Kriterium kann eine nicht zu überschreitende Differenz in den *propensity scores* von geförderten und nicht geförderten Personen festgelegt werden. Bei „Ziehen mit Zurücklegen“ kann dieselbe Vergleichsperson unterschiedlichen geförderten Personen zugeordnet werden.

⁴ Diese Annahme lässt sich formal darstellen als: $(Y_0, Y_1) \perp (D|X)$, wobei Y_0 und Y_1 die Ergebnisvariablen bei Nichtteilnahme bzw. Teilnahme sind, \perp für Unabhängigkeit stehen soll, die Variable D mit den Werten 0 und 1 Nichtteilnahme bzw. Teilnahme bezeichnet und X der Vektor der Variablen ist, auf die mit dem *matching* konditioniert wird. Soll der durchschnittliche Fördererfolg für die geförderten Personen (*mean effect of treatment on the treated*) geschätzt werden, genügt die Annahme: $Y_0 \perp (D|X)$, sie wird auch als *conditional independence assumption* bezeichnet.

levanten Merkmale von Teilnehmer/innen- und Vergleichsgruppe betrifft nicht nur *matching*-Verfahren, wird hier aber deutlicher als bei anderen Evaluationstechniken.

Als Vorteil des *matching*-Verfahrens kann gelten, dass es sich hierbei um eine nichtparametrische Technik handelt. Weder muss angenommen werden, dass das Förderergebnis im Personenvergleich identisch ist, noch muss für den Zusammenhang zwischen Personenmerkmalen und Förderergebnissen eine funktionale Form unterstellt werden.

Die Güte einer auf der Bildung von Vergleichsgruppen basierenden Evaluation wird stets davon abhängig sein, inwieweit es gelingt, mit den beobachteten Merkmalen „Vergleichbarkeit“ zwischen den geförderten Personen und den hinzugespielten Vergleichspersonen herzustellen. Dies stellt hohe Anforderungen an die Datenqualität und stößt an Grenzen dort, wo die Überlappung der Zugangswahrscheinlichkeiten bzw. der Merkmale von Teilnehmer/innen und Vergleichsgruppe auch in der Realität nicht gegeben ist. Wenn nicht zusätzliche Annahmen getroffen werden, muss die Schätzung des Förderergebnisses auf den Überlappungsbereich beschränkt bleiben.

Schließlich sei hier noch ein bisher ausgeblendetes Problem angesprochen, das nicht nur für Anwendungen von *matching*-Techniken gilt, sondern generell für mikroökonomische Wirkungsanalysen auftritt, wenn die Vergleichsgruppenbildung als Weg dient, die kontrafaktische Situation der geförderten Personen nachzubilden. So wird regelmäßig die Annahme getroffen, dass das untersuchte Förderinstrument keine Wirkungen auf die nicht geförderten Personen ausübt. Das Auftreten derartiger indirekter Wirkungen ist aber, besonders wenn es sich um große Förderinstrumente handelt, wahrscheinlich. Ein wichtiges Beispiel sind Verdrängungseffekte, durch die nicht geförderte Personen oder nicht geförderte Betriebe in ihren Wettbewerbschancen beeinträchtigt werden. Auch der mit der Förderung entstehende Finanzierungsbedarf kann zu negativen Entzugseffekten führen, die nicht unbedingt gleichermaßen für geförderte und nicht geförderte Personen gelten müssen. Ebenso denkbar sind positive indirekte Wirkungen, wenn etwa die regionale Konzentration von Fördermitteln zu einer Besserung der lokalen Arbeitsmarktentwicklung insgesamt beiträgt.

4 Vergleichsgruppenbildung und Analyse von Zugangsprozessen an einem Beispiel

Im Rahmen eines laufenden IAB-Projekts⁵ werden die individuellen Beschäftigungswirkungen betrieblicher Einstellungshilfen, das sind in diesem Fall die Lohnkostenzuschüsse an Arbeitgeber nach SGB III und die Beschäftigungshilfen für Langzeitarbeitslose, unter-

⁵ [IAB-Projekt 10-497](#): „Selektivität und Eingliederungserfolg betriebsbezogener Einstellungshilfen“.

sucht. Im Vordergrund stehen hierbei nicht Fragen nach der Dauer und der Stabilität der geförderten Beschäftigung und nachfolgenden ungeförderten Beschäftigungsverhältnissen. Stattdessen soll untersucht werden, in welchem Umfang die Förderung tatsächlich zu einer Verbesserung der Beschäftigungschancen bzw. zu einer Verringerung des Arbeitslosigkeitsrisikos der geförderten Personen beigetragen hat. Da dies auch die Frage beinhaltet, welche Beschäftigungschancen diese Personen ohne Förderung oder bei erst später einsetzender Förderung gehabt hätten, ist das Heranziehen einer Vergleichsgruppe nicht geförderter Personen ein wichtiger Projektbestandteil.

Eine Besonderheit des Projekts ist, dass die wichtigsten bestehenden Formen von Lohnkostenzuschüssen gemeinsam evaluiert werden. Hierdurch werden Analysen der Förderwirkungen im Personen- und Instrumentenvergleich ermöglicht. Weiterhin wurde besondere Sorgfalt auf die Erfassung individueller Merkmale mit potenziellem Einfluss auf die Wahrscheinlichkeit des Förderzugangs gelegt. Die Stichprobe der geförderten Personen setzt sich aus in den ersten Monaten des Jahres 1999 in ausgewählten Arbeitsamtsbezirken erfolgten Zugängen in die Förderung zusammen. Parallel wurden aus den coArb-Daten der ausgewählten Ämter zur Bildung der Vergleichsgruppe Zufallsstichproben aus im Januar 1999 arbeitslos gemeldeten Personen gezogen. Für die geförderten Personen und für die Personen der Vergleichsstichprobe wurde eine schriftliche Zusatzerhebung bei den jeweils zuständigen Arbeitsvermittler/innen durchgeführt. Neben Angaben zur Erwerbsbiographie, zur regionalen Mobilität und zu qualifikatorischen Defiziten wurden auch sensible Personeninformationen erfragt, so z. B. Einschätzungen der Vermittler/innen zur Motivation und zu den Suchanstrengungen während der Arbeitslosigkeit. Hierdurch stehen auch Personenmerkmale, die sonst eher mit dem Begriff „unbeobachtete Heterogenität“ umrissen werden, für die Untersuchung möglicher Selektivität beim Förderzugang zur Verfügung.

Ein Vorteil dieses Projektdesigns ist, dass durch die zeitgleiche Ziehung der Stichproben geförderter Personen und nicht geförderter Personen beide Gruppen derselben Population entstammen. Ferner wurden mit diesem Design die zugangsrelevanten Personenmerkmale zeitnah erhoben. Da die Förderwirkung aber durch den Verbleib in ungeförderter Beschäftigung geschätzt werden soll, kann der sehr lange Untersuchungszeitraum, bedingt durch Förderdauern, die häufig bei einem Jahr oder darüber liegen, als nachteilig angesehen werden.

Eine erste Untersuchung der Zugangsprozesse (vgl. Jaenichen 2000) zeigte, dass hinsichtlich der Struktur der geförderten Personen vor allem Unterschiede zwischen Förderarten mit weniger restriktiven Fördervoraussetzungen (Eingliederungszuschüsse bei Einarbeit, Einstellungszuschüsse bei Neugründungen, Strukturanpassungsmaßnahmen Ost für Wirtschaftsunternehmen) und stärker auf Schwervermittelbarkeit bzw. Langzeit-

arbeitslosigkeit ausgerichteten Förderarten (Eingliederungszuschüsse bei erschwerter Vermittlung, Beschäftigungshilfen für Langzeitarbeitslose) bestehen. Im Vergleich mit den ohne Förderung in Beschäftigung gewechselten Personen der Vergleichsstichprobe zeigten sich bei einer Simulation potenzieller Arbeitslosigkeitsdauern wiederum eher geringe Unterschiede für die durch Instrumente mit schwachen Fördervoraussetzungen geförderten Personen, stärkere Unterschiede dagegen für die Personen in stärker auf Schwervermittelbarkeit abzielenden Förderarten. Im Vergleich zu den nicht in reguläre Beschäftigung gewechselten Personen der Vergleichsstichprobe wiesen jedoch auch die durch stärker auf Schwervermittelbarkeit abzielenden Lohnkostenzuschüsse geförderten Personen eher niedrige geschätzte Risiken des Verbleibs in Arbeitslosigkeit auf.

Die Ergebnisse dieser Zugangsuntersuchung deuten somit darauf hin, dass durchaus Selektivität bei der Förderung durch Lohnkostenzuschüsse gegeben ist. Ferner ist zu vermuten, dass von der Mehrzahl der den Förderzugang beeinflussenden Merkmalen auch Einflüsse auf die Beschäftigungschancen ausgehen. Bei der noch durchzuführenden Evaluation des Eingliederungserfolgs für die geförderten Personen mittels Vergleichsgruppenbildung werden daher die für eine jeweilige Personengruppe oder für ein jeweiliges Förderinstrument gegebenen Zugangsprozesse zu berücksichtigen sein.

5 Schlussbemerkungen

Mit dem vorliegenden Beitrag sollten ohne explizite Darstellung des formalen Rahmens grundlegende Aspekte der Vergleichsgruppenbildung bei der mikroökonomischen Evaluation arbeitsmarktpolitischer Instrumente erläutert werden. Im Zentrum mikroökonomischer Evaluationsmethoden steht die Behandlung des *selection bias*, der möglichen Verzerrung des geschätzten Fördererfolgs durch systematische, eventuell aber nicht beobachtete Unterschiede zwischen geförderten Personen und der Gruppe potenzieller Vergleichspersonen. Auch die experimentelle Zufallsauswahl kann dieses Problem nur teilweise lösen. Ein wesentlicher Vorteil einer experimentellen Datenbasis besteht darin, dass Personen zufällig von der Förderung ausgeschlossen werden und daher auch im Bereich sehr hoher Förderwahrscheinlichkeiten prinzipiell Vergleichspersonen existieren. *Matching*-Techniken imitieren Experimente, indem die Vergleichsgruppe durch Konditionierung auf die die Selektionsverzerrung verursachenden Merkmale bestimmt wird. Voraussetzung ist hier allerdings, dass diese Merkmale auch beobachtet werden. Ist dies nicht der Fall, müssen zusätzliche Annahmen für eine Abschätzung des Fördererfolgs getroffen werden. Für Alternativen zu *matching*-Verfahren sei hier auf die umfassende Darstellung von Heckman u. a. (1999) oder auch auf den Überblicksartikel von Fitzenberger/Speckesser (2000) verwiesen.

Als Beispiel für ein auf Vergleichsgruppenbildung basierendes Projektdesign wurde hier das IAB-Projekt zu individuellen Beschäftigungswirkungen betrieblicher Einstellungshilfen kurz vorgestellt. Auch andere neue IAB-Projekte arbeiten mit dem Vergleichsgruppenkonzept.⁶ Hervorzuheben sind dabei die Anstrengungen, die unternommen werden, um Prozessdaten aus dem Geschäftsbereich der BA (Beschäftigtenstatistik, Leistungsempfängerdatei, frühere FuU-Statistik, aktuelle Förder- und Bewerberangebotsdateien) für Evaluationszwecke zu erschließen.

Literatur

- Ashenfelter, O. (1978): Estimating the effect of training programs on earnings, *Review of Economics and Statistics* 6 (1), 47-57.
- Brinkmann, Ch./Wießner, F. (2002): [Zur Wirkungsforschung aktiver Arbeitsmarktpolitik - neue Herausforderungen für Monitoring und Evaluation](#).
- Fitzenberger, B./Speckesser, S. (2000): Zur wissenschaftlichen Evaluation der Aktiven Arbeitsmarktpolitik in Deutschland: Ein Überblick, *MittAB* 33 (3), 357-370.
- Ham, J.C./LaLonde, R.J. (1990): Using social experiments to estimate the effect of training on transition rates, in: Hartog, J./Ridder, G./Theeuwes, J. (Hrsg.): *Panel data and labor market studies*, Elsevier Science Publishers.
- Heckman, J.J./Ichimura, H./Todd, P.E. (1997): Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme, *Review of Economic Studies* 64, 605-654.
- Heckman, J.J./LaLonde, R.J./Smith, J.A. (1999): The Economics and Econometrics of Active Labor Market Programs, in: Ashenfelter, O./Card, D. (Hrsg.): *Handbook of Labor Economics*, Vol. IIIA, 1865-2097.
- Jaenichen, U. (2000): Selektivität beim Zugang in Förderung durch betriebliche Einstellungshilfen, *MittAB* 33, 445-461.
- Lechner, M./Gerfin, M. (2000): Wirkung der aktiven Arbeitsmarktpolitik der Schweiz auf die individuellen Beschäftigungschancen von Arbeitslosen, *MittAB* 33 (3), 396-404.
- Rosenbaum, P.R./Rubin, D.B. (1983): The central role of the propensity score in observational studies for causal effects, *Biometrika* 70(1), 41-55.
- Sianesi, B. (2001): An evaluation of the active labour market programmes in Sweden, IFAU working paper 2001:5.

⁶ Vgl. „Brinkmann/Wießner: [Zur Wirkungsforschung aktiver Arbeitsmarktpolitik](#)“.