ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Schmidt, Christoph M.

Working Paper

Arbeitsmarktpolitische Maßnahmen und ihre Evaluierung: eine Bestandsaufnahme

IZA Discussion Papers, No. 207

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Schmidt, Christoph M. (2000) : Arbeitsmarktpolitische Maßnahmen und ihre Evaluierung: eine Bestandsaufnahme, IZA Discussion Papers, No. 207, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/21052

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 207

Arbeitsmarktpolitische Maßnahmen und ihre Evaluierung: eine Bestandsaufnahme

Christoph M. Schmidt

October 2000

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Arbeitsmarktpolitische Maßnahmen und ihre Evaluierung: eine Bestandsaufnahme

Christoph M. Schmidt
University of Heidelberg, CEPR, London and IZA, Bonn

Discussion Paper No. 207
October 2000

IZA

P.O. Box 7240
D-53072 Bonn
Germany

Tel.: +49-228-3894-0
Fax: +49-228-3894-210
Email: iza@iza.org

This Discussion Paper is issued within the framework of IZA's research area *Project Evaluation.* Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent, nonprofit limited liability company (Gesellschaft mit beschränkter Haftung) supported by the Deutsche Post AG. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public. The current research program deals with (1) mobility and flexibility of labor markets, (2) internationalization of labor markets and European integration, (3) the welfare state and labor markets, (4) labor markets in transition, (5) the future of work, (6) project evaluation and (7) general labor economics.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character.

IZA Discussion Paper No. 207
October 2000

ABSTRACT

Arbeitsmarktpolitische Maßnahmen und ihre Evaluierung: eine Bestandsaufnahme*

(Title: Active Labor Market Policy and its Evaluation: The Current State of Affairs)
Despite the considerable effort and the large financial expenditures spent on measures of Active Labor Market Policy in Germany, relatively little is known about their actual impact. To the contrary, past evaluation efforts typically failed to construct a credible counterfactual situation – a precise statement of what would have happened in the absence of the policy intervention. This paper argues that Germany lags clearly behind the current evaluation practice of other advanced economies, and in information about the efficacy of its policies. The paper discusses the methodological problems surrounding the evaluation of policy interventions and possible empirical strategies for their solution. The available scientific evidence on the impact of labor market programs is surveyed briefly, and several lessons are extracted on the design of future programs and their evaluation.

Zusammenfassung:
Trotz des beachtlichen personellen und finanziellen Aufwands, der in der Bundesrepublik für die Durchführung arbeitsmarktpolitischer Maßnahmen betrieben wird, ist ihr Erfolg in keiner Weise gesichert. Im Gegenteil, die bisherige Evaluierungspraxis verletzt nahezu durchgängig eine der Minimalanforderungen an zielgerichtetes und erfolgsorientiertes Handeln, die Erfordernis, eine Vergleichssituation zu konstruieren, die hinreichend genau beschreibt, was sich ohne den in Frage gestellten Eingriff ergeben hätte. Vor dem Hintergrund eines internationalen Vergleichs der Evaluierungspraxis diskutiert dieser Beitrag die methodischen Probleme bei der Evaluierung arbeitsmarktpolitischer Maßnahmen, wobei unterschiedliche empirische Strategien zur Lösung dieser Probleme vorgestellt werden. Darüber hinaus wird ein kurzer Abriß der verfügbaren wissenschaftlichen Evidenz zu ökonomischen Auswirkungen von Arbeitsmarktprogrammen angeboten, aus denen eine Reihe zentraler Schlußfolgerungen für zukünftige Vorschläge solcher Maßnahmen und ihre Evaluierung gezogen wird.

JEL Classification: J68, H43, H53

Keywords: Aktive Arbeitsmarktpolitik, Qualifikation, Ausbildung, Humankapital, Subventionen, ABM, SGB III, Empirische Strategie, Kontrafaktische Situation, Identifikation, Experiment

Christoph M. Schmidt
Department of Economics (Econometrics)
University of Heidelberg
Grabengasse 14
69177 Heidelberg
Germany
Tel.: +49-6221-54 2955
Fax: +49 6221 54 3640
Email: schmidt@uni-hd.de

* Ich danke Boris Augurzky, Jochen Kluve und Ralph Würthwein für ihre Kommentare.

1 Arbeitsmarktpolitische Maßnahmen

Ein zentrales Problem vieler entwickelter Volkswirtschaften ist die hohe Arbeitslosigkeit, die angesichts ihres Beharrungsvermögens offenbar nicht befriedigend als die Auswirkung bloßer konjunktureller Schwankungen um ein gemäßigtes Beschäftigungswachstum erklärt werden kann. Ökonomen argumentieren stattdessen häufig, daß hier eine zu geringe *Lohnspreizung* und eine geringe berufliche *Qualifikation* und Qualifizierbarkeit der Betroffenen ineinandergreifen, vor allem da Arbeitslosigkeit überwiegend Personen mit geringer oder nicht zeitgemäßer Ausbildung trifft.

Während auf der Seite der Arbeitgeber beklagt wird, daß nicht in hinreichendem Maße Arbeitnehmer zu finden sind, die bereit sind, niedrig qualifizierte und somit auch niedrig entlohnte Arbeiten anzunehmen, bzw. daß es nicht möglich ist, ohne die Hemmnisse von Flächentarifverträgen und ohne bürokratische Widrigkeiten neue Stellen im Niedriglohnbereich zu schaffen, fällt es Arbeitslosen häufig schwer, ihre Ausbildungsdefizite durch weitere Bildungsinvestitionen wettzumachen. Zudem sind die Rahmenbedingungen des Arbeitsmarkts der Entwicklung von Eigeninitiative wenig förderlich. So wird von allen Seiten bemängelt, daß der deutsche Arbeitsmarkt an zu wenig Flexibilität und Mobilität krankt. Zur Qualifikations- und Altersstruktur von Arbeitslosigkeitsraten und Abgangsraten aus der Arbeitslosigkeit siehe z.B. Schmidt (2000a).

Mögliche wirtschaftspolitische Maßnahmen zur Bekämpfung der Arbeitslosigkeit lassen sich in zwei Ansätze untergliedern. Der erste Ansatz betrifft die Umgestaltung des Systems der sozialen Sicherung und andere, flankierende Reformen zur Förderung der wirtschaftlichen Rahmenbedingungen mit dem Ziel, diese so neu zu gestalten, daß Motivation, Kreativität und Leistungswillen freigesetzt werden können. Ein zweiter Ansatz zielt direkt auf die Problemgruppen der niedrig qualifizierten und nicht zeitgemäß qualifizierten Arbeitnehmer ab, ohne das System der sozialen Sicherung insgesamt zu verändern. Seit Jahrzehnten sind die öffentliche Bereitstellung von Ausbildungsmaßnahmen, Hilfe bei der Arbeitsplatzsuche, Lohnsubventionen für Arbeitnehmer oder Arbeitgeber sowie Arbeitsbeschaffung im öffentlichen Sektor Teile der sozialen Wohlfahrtspolitik in allen entwickelten Marktwirtschaften.

So sieht auch die Beschäftigungsförderung in der Bundesrepublik erhebliche finanzielle Leistungen vor. Als Maßnahmen sog. *Aktiver Arbeitsmarktpolitik* bezeichnet man die Gesamtheit der Aktivitäten der Arbeitsministerien von Bund und Ländern und der diesen angeschlossenen Verwaltungsapparate -- insbesondere der Bundesanstalt für Arbeit --, bei deren Einsatz ein Ermessensspielraum besteht. Dazu zählen u. a. Umschulungs- und Fortbildungsprogramme zur Förderung und zum Erhalt des Kenntnisstands Arbeitsloser oder von Arbeitslosigkeit bedrohter

Arbeitnehmer, das Design von Anreizschemata zur Beeinflussung des Arbeitsangebotsverhaltens von Sozialhilfeempfängern, aber auch die direkte Bereitstellung neuer Arbeitsplätze im Rahmen von Arbeitsbeschaffungsmaßnahmen. Diese drei Arten von Maßnahmen, (i) die *Förderung von Humankapital*, (ii) die *Veränderung von Anreizen* und (iii) die *Bereitstellung von Arbeitsplätzen* stellen, unabhängig von ihrer konkreten Implementierung -- so können Arbeitsplätze beispielsweise von öffentlichen Unternehmen oder durch Subventionierung der Lohnkosten privater Unternehmen bereitgestellt werden -- die drei Grundformen aktiver Arbeitsmarktpolitik dar.

Während in den Alten Bundesländern die Förderung von Humankapital, eine im Einzelfall durchaus kostenintensive Maßnahme, anhand der Zahl der Betroffenen wie auch der Ausgabeanteile mit etwa zwei Dritteln des Etats im Vordergrund steht, wird in den Neuen Bundesländern für Maßnahmen direkter Arbeitsbeschaffung, ebenfalls kostenintensiv in der pro-Kopf-Berechnung, nahezu die Hälfte des verfügbaren Budgets verausgabt. Verwaltungstechnisch sind es eben jene Haushaltsmittel der Bundesanstalt für Arbeit für die Bereitstellung von *Ermessensleistungen*, die den Gesamtumfang der Maßnahmen aktiver Arbeitsmarktpolitik entscheidend bestimmen. Seit der Reform der rechtlichen Rahmenbedingungen der Arbeitsförderung (SGB III) des Jahres 1998 wird über die konkrete Vergabe dieser Mittel -- und somit ihre Aufteilung auf die drei Grundformen aktiver Arbeitsmarktförderung, im Jahre 1998 insgesamt rund 25 Mrd. DM, dezentral auf der Ebene der Arbeitsamtsbezirke entschieden. Somit ergibt sich eine gewisse Variation des Maßnahmenbündels über einzelne Regionen hinweg, auch wenn der West-Ost-Unterschied diese Variation klar dominiert.

Nicht zur Aktiven Arbeitsmarktpolitik zählen dagegen Ausgaben für Arbeitslosengeld bzw. -hilfe und alle anderen Maßnahmen, auf die Arbeitnehmer, Arbeitslose oder auch Berechtigte, die außerhalb des Arbeitsmarktes stehen, einen rechtlichen Anspruch besitzen (im Jahre 1998 nahezu 90 Mrd. DM). Bei diesen Maßnahmen besitzen die Organe der Arbeitsförderung somit keinerlei Ermessensspielraum. Es fehlt ihnen somit jegliches aktive Steuerungselement. In ihrer Gesamtheit bezeichnet man sie daher auch als Eingriffe der *Passiven Arbeitsmarktpolitik*. Sowohl zur inhaltlichen Abgrenzung der unterschiedlichen Maßnahmen als auch der verwaltungstechnischen Einzelheiten des SGB III vgl. Fertig und Schmidt (2000).

Eines jedoch eint Maßnahmen Aktiver und Passiver Arbeitsmarktpolitik, die hohen Kosten, mit denen sie die öffentlichen Haushalte belasten. Welchen Standpunkt auch immer man einnehmen mag, was die angemessene Rolle des Staates anbetrifft -- eine Debatte, die die Nationalökonomie seit Jahrhunderten beschäftigt, mit Idealvorstellungen, die vom nahezu vollkommen zurückhaltenden Staat bis hin zum modernen Wohlfahrtsstaat kontinentaleuropäischer bzw. nordischer Länder reichen -- was für arbeitsmarktpolitische Maßnahmen verausgabt wird,

steht bei gegebenem Budget nunmal nicht mehr für andere staatliche Aktivitäten zu Verfügung. Auch die Größe dieses Budgets ist ja keineswegs sakrosankt. Geringere Ausgaben des Staates würden auch ein geringeres Einnahmevolumen verlangen, sprich geringe Steuern. Das heißt natürlich, daß ein größerer Anteil des erwirtschafteten Einkommens bei den privaten Haushalten verbleiben könnte - möglicherweise ein eigener, wichtiger Wachstumsmotor.

Man muß sich also schon viel davon versprechen, wenn man solch substantielle Mittel dazu verwendet, um das Problem der Arbeitslosigkeit zu bekämpfen. Der Erfolg dieser Maßnahmen ist allerdings keineswegs gesichert. Im Gegenteil, um erfolgreich zu sein, muß der Ursache des Problems entgegengewirkt werden -- fehlen Arbeitslosen beispielsweise die nötigen Fertigkeiten, so wird ihre Beschäftigung im Rahmen einer Arbeitsbeschaffungsmaßnahme wohl eher nicht dazu geeignet sein, sie erfolgreich in den sog. Ersten Arbeitsmarkt zurückzuführen. Die Ursachen der bestehenden Arbeitslosigkeit sind jedoch mannigfaltig und variieren von Zielgruppe zu Zielgruppe, über Regionen und Zeitpunkte hinweg. Daher ist immer fraglich, ob in der gegebenen Anwendungssituation auch tatsächlich das angemessene Maßnahmenbündel gewählt worden ist. Darüber hinaus reicht es nicht aus, lediglich eine gewisse Wirkung zu erzielen. Vielmehr erfordern die typischerweise hohen Kosten arbeitsmarktpolitischer Eingriffe auch Wirkungen beträchtlichen Ausmaßes, soll ihr Einsatz gesellschaftlich gerechtfertigt sein[1].

Somit erwartet man völlig zu recht, daß arbeitsmarktpolitische Maßnahmen zielgerichtet und erfolgsorientiert durchgeführt werden. Allerdings kann dies im Augenblick eigentlich überhaupt nicht der Fall sein, so hart ein solch apodiktisches Urteil auch wirken mag. Ein Satz von Minimalanforderungen an zielgerichtetes und erfolgsorientiertes Handeln -- Allgemeingut jedes selbständigen Unternehmers, jedes privaten Geldanlegers, aber natürlich auch jedes Haushaltsvorstandes -- ist die Vorgabe eines Ziels, das erreicht werden soll (hier z.B. Verbesserung der Arbeitsmarktsituation niedrig ausgebildeter Arbeitnehmer), die Definition eines Erfolgskriteriums (hier z.B. eine kurze durchschnittliche Verweildauer der Arbeitslosen in der Beschäftigungslosigkeit) und, was besonders wichtig ist, die Konstruktion einer Vergleichssituation, die hinreichend genau beschreibt, was sich ohne den in Frage gestellten Eingriff entwickelt hätte. In technischen Termini ist dies die sog. *kontrafaktische* Situation. Wie im folgenden ausgeführt werden wird, ist die Konstruktion des Kontrafaktischen in der Regel eine hohe intellektuelle Herausforderung; unglücklicherweise sind verwaltungstechnische und

[1] Bei manchen Zielgruppen, beispielsweise benachteiligten Jugendlichen, mag auch ein ungünstigeres Verhältnis von Ertrag und Kosten noch tolerabel sein. Diese Gewichtung ist Teil des politischen Prozesses und kann hier nicht näher diskutiert werden. Was aber wichtig ist, ist die nach dem heutigen Stand der Erkenntnisse bestmögliche Abschätzung der Effekte und Kosten gegenüber dem Steuerzahler offen zu legen.

betriebswirtschaftliche Steuerungsinstrumente wie beispielsweise das so magisch klingende „Controlling öffentlicher Unternehmen“' dazu in der Regel gänzlich ungeeignet.

Dieser Beitrag skizziert vor dem Hintergrund der quantitativen Bedeutung arbeitsmarktpolitischer Eingriffe den Stand der „Evaluierungskultur“ bei der Bewertung dieser Maßnahmen im internationalen Vergleich (Abschnitt 2), bietet eine komprimierte Darstellung der methodischen Probleme und des zu ihrer Lösung verfügbaren wissenschaftlichen Instrumentariums (Abschnitt 3), faßt die verfügbare internationale Evidenz zu ökonomischen Auswirkungen von Arbeitsmarktprogrammen zusammen (Abschnitt 4), wobei aus konzeptionellen Gründen die Betonung auf der umfangreichen nordamerikanischen Erfahrung mit Arbeitsmarktmaßnahmen und ihrer wissenschaftlichen Evaluierung liegen muß, und zieht im letzten Abschnitt eine Reihe zentraler inhaltlicher Schlußfolgerungen für zukünftige Vorschläge arbeitsmarktpolitischer Maßnahmen und ihre wissenschaftliche Evaluation.

2 Der Stand internationaler Evaluierungskultur

Bevor im folgenden Abschnitt näher darauf eingegangen werden soll, wie das sog. Evaluierungsproblem genauer zu charakterisieren ist und welche Lösungsmöglichkeiten die Wissenschaft anzubieten vermag, soll hier ein kurzer Abriß des Stands der Evaluierungskultur im internationalen Vergleich angeboten werden. Insbesondere wird hier die Frage diskutiert, warum diese Diskussion für die Bundesrepublik erst jetzt ein hinreichendes Gewicht erhält. So herrscht zwar auch im nordamerikanischen Kontext durchaus noch punktueller Verbesserungsbedarf, Arbeitsökonomen war aber schon geraume Zeit das deutsche bzw. europäische Evaluierungsdefizit bei staatlichen Eingriffen jeder Art offenbar.

Ein kurzer Blick über den Atlantik reichte aus, um die Unterschiede im Ansatz zu Tage treten zu lassen: ein in der nordamerikanischen Gesellschaft fest verwurzeltes Mißtrauen gegenüber aller staatlicher Lenkung, in Kombination mit der Betonung eigenverantwortlichen Handelns, ließe gar nicht erst zu, arbeitsmarktpolitische Aktivitäten ohne die Bewertung durch unabhängige Gutachter, also aufgrund der wissenschaftliches Instrumentarium verlangenden (siehe Abschnitt 3) Natur des Problems unabhängige Wissenschaftler, durchzuführen. Im Gegenteil, das Recht der Öffentlichkeit, so genau wie möglich und durch institutionell unabhängige Gutachter über Kosten und Ertrag staatlichen Handelns informiert zu werden, erscheint in Nordamerika wie ein

Grundrecht[2]. Daher werden in den USA auch seit langem beträchtliche Mittel für die wissenschaftliche Evaluation von Arbeitsmarktprogrammen verausgabt.

Wie fern ist ein solches Vorgehen den kontinentaleuropäischen Gesellschaften! Sind es doch keineswegs Privatfirmen, die im Auftrag des Staates hoheitliche Maßnahmen umsetzen, sondern konstituierende Elemente des meritorischen Staats höchstselbst. Auch wenn man sich aus heutiger Sicht durchaus fragen mag, warum beispielsweise die Beförderung von Briefen eine hoheitliche Aufgabe[3] oder warum der Lateinlehrer der Tochter verbeamtet sein muß, so ist die Art und Weise, mit der in Europa diese Aufgaben wahrgenommen werden, doch auch hochbewährt. Insbesondere ist ihre Delegation an einen Beamtenapparat oft mit der Erwartung verknüpft, daß sie mit hoher Qualität erfüllt werden. Daher muß man sich schon fragen, warum man sich davor scheuen sollte, die Frage nach der tatsächlichen Wirkung staatlicher Maßnahmen offen zu stellen[4].

Hinzu kommt, daß, im Gegensatz zum nordamerikanischen Erfahrungsschatz Maßnahmen Aktiver Arbeitsmarktpolitik als Instrumente zur Bekämpfung von Arbeitslosigkeit und Armut in Europa ein relativ junges Phänomen darstellen. In den Ökonomien Westeuropas bedurfte es für lange Zeit keiner solchen Maßnahmen, da die Ära nach dem Zweiten Weltkrieg in ganz Westeuropa eine Phase ausserordentlich hohen ökonomischen Wachstums darstellte. Erst die dauerhaft auf hohem Niveau verharrenden Arbeitslosigkeitsraten der 80er und 90er Jahre des vergangenen Jahrhunderts haben Arbeitsmarktprogramme an Bedeutung gewinnen lassen. Ähnlich ging es den früheren Mitgliedsstaaten des Ostblocks, die jetzt das Spektrum der Übergangsökonomien auf dem Weg zu modernen Marktwirtschaften bilden. Jetzt, da einige von ihnen vor dem Eintritt in die *Europäische Union* stehen, müssen sie zum ersten Mal das Phänomen umfangreicher *offener* Arbeitslosigkeit bekämpfen, das früher schlichtweg wegdefiniert war. Von Beginn ihrer Transformation an haben die Übergangsökonomien Zentral- und Osteuropas große Teile des Aufbaus der westlichen Sozialversicherungsysteme übernommen -- ohne daß eigentlich hinreichend Kenntnis über deren Wirksamkeit vorhanden sein konnte.

Natürlich wurden seit Beginn der Einführung von Arbeitsmarktmaßnahmen in das Spektrum staatlicher Interventionen in Europa Anstrengungen unternommen, ihre Wirkungen wissenschaftlich zu durchleuchten. Nichtsdestoweniger zeigt ein sorgfältiger Überblick über den bestehenden Erfahrungsschatz, daß diese Forschung noch in den Anfängen steckt. Vor allem das

[2] Daß sich im Laufe der letzten Jahrzehnte einige privatwirtschaftliche Firmen auf dieses „Evaluierungsgeschäft“ spezialisiert haben, wirkt dort nicht im geringsten als störend. Sind es doch häufig privatwirtschaftlich organisierte Firmen, nicht Beamte, die staatliche Maßnahmen implementieren.

[3] Die anstehenden rechtlichen Änderungen in diesem Bereich sind dem Autor durchaus bewußt.

vorhandene Datenmaterial und das verfügbare Studiendesign können oft nicht die Ansprüche erfüllen, die von fortgeschrittenen ökonometrischen Methoden gestellt werden. Am markantesten ist die nahezu vollkommene Abwesenheit sozialwissenschaftlicher Experimente in Europa -- dabei waren es in der Vergangenheit gerade sozialwissenschaftliche Experimente, die einen großen Teil der verlässlichsten und miteinander im Einklang stehenden nordamerikanischen Evidenz zur Wirksamkeit von Arbeitsmarktprogrammen erzeugen konnten.

Die von Wissenschaftlern häufig gestellte Forderung nach mehr sozialwissenschaftlichen Experimenten hat ihre Wurzeln in den praktischen Erfahrungen mit der Evaluierung der wichtigsten nordamerikanischen Ausbildungsprogramme (siehe Schmidt (2000b)). Die umfangreiche nicht-experimentelle Evidenz, die zu den zwei größten während der 60er und 70er Jahre des vergangenen Jahrhunderts bundesweit implementierten Ausbildungsprogrammen der USA, dem *Manpower Development and Training Act* (MDTA) und dem *Comprehensive Employment and Training Act* (CETA), gesammelt wurde, offenbarte eindeutig, daß die spezifische Modellierungsstrategie die Ergebnisse jeder Evaluierungsstudie entscheidend mitprägte. Dabei führten unterschiedliche, aber unter Verhaltensgesichtspunkten gleichermaßen ansprechende *Identifikationsannahmen* (siehe auch Abschnitt 3) oft zu weit auseinanderliegenden Ergebnissen. Im Laufe der 80er Jahre verfestigte sich der Eindruck, daß auch eine weitere Verfeinerung der statistischen und ökonometrischen Methoden an dieser Variabilität nichts mehr ändern könnte, da sie weitgehend auf die Natur nicht-experimenteller Evaluierungsstrategien zurückzuführen sei.

Unter diesem Eindruck entschloß man sich, das wichtigste Ausbildungsprogramm der 80er Jahre, den *Job Training Partnership Act* (JTPA) mit einem groß angelegten sozialwissenschaftlichen Experiment zu evaluieren. Nichtsdestoweniger wurden parallel die nicht-experimentellen Techniken weiterentwickelt. In der jüngeren Literatur zu nicht-experimentellen Evaluierungsstudien zeigt sich, daß viele der Probleme, die die Ergebnisse der älteren Generation nicht-experimenteller Studien noch unvermeidlich verzerrt hatten, mittlerweile weitgehend gelöst sind. Diese Entwicklung wird allerdings nicht nur durch verbesserte statistische Methoden, sondern auch durch ein verbessertes Datenmaterial erzielt. In jedem Falle behalten sozialwissenschaftliche Experimente ihre dominante konzeptionelle Rolle als die Evaluationsstrategie, die letztlich in der Lage ist, die glaubwürdigsten und stabilsten Ergebnisse zu erzielen, eine Erkenntnis, die den Weg in die bundesdeutsche „Evaluierungskultur“ zum gegenwärtigen Zeitpunkt noch finden muß.

[4] Es ensteht natürlich ein Problem des Managements öffentlicher Einrichtungen: wie bei jeder Bewertung stellt alleine der Akt des Hinterfragens der Leistung einen ernsten Angriff auf die Leistungsbereitschaft der Bewerteten dar, wenn sie nicht von breiter Akzeptanz getragen oder von ernsthaften persönlichen Konsequenzen begleitet wird.

3 Methodische Aspekte

3. 1 Zentrale Elemente der Evaluierung staatlicher Eingriffe

Unabhängig von der konkreten Natur der arbeitsmarktpolitischen Maßnahme erfordert jeglicher Versuch einer Evaluierung, daß für wohldefinierte Beobachtungseinheiten -- je nach Eingriff Arbeitnehmer oder Betriebe, aber auch ganze Regionen -- die zu beobachtenden relevanten Zielgrößen nach der Durchführung der Maßnahme, beispielsweise Beschäftigungserfolg oder Löhne, Beschäftigungszuwachs des Betriebs oder die Arbeitlosigkeitsrate der Region, mit den Ergebnissen einer Vergleichssituation konfrontiert werden (siehe z.B. Heckman et al. (1999), Manski (1995) oder Schmidt (1999)). Welche Beobachtungseinheiten und Zielgrößen in einer solchen Analyse relevant sind, wird maßgeblich durch die Art der Maßnahme und der Daten bestimmt, die für die Analyse zur Verfügung stehen. So wird eine Lohnsubvention, die direkt an Arbeitnehmer ausgezahlt wird, vorzugsweise mittels einer Analyse des Arbeitsmarktverhaltens und des -erfolgs individueller Arbeitnehmer untersucht[5].

In ähnlicher Form verläuft die Suche nach einem angemessenen Erfolgskriterium. Ist die Bereitstellung einer Lohnsubvention an Arbeitnehmer erfolgreich, so sollte sie bei den Betroffenen zu einer Verbesserung der Arbeitsmarktsituation führen (wobei die Erfolgreichen diese Subvention dann abrufen würden). Während man bei der Evaluierung arbeitsmarktpolitischer Programme im nordamerikanischen Kontext häufig an höhere Löhne denkt (dauerhafte Arbeitslosigkeit spielt in diesem Arbeitsmarkt eine weniger wichtige Rolle als in europäischen Arbeitsmärkten), so stünde bei der Bewertung der Wirkung einer Lohnsubvention in Deutschland wohl eher eine höhere Beschäftigungswahrscheinlichkeit im Zentrum des Interesses. Die Wahl des geeigneten Erfolgskriteriums ist jedoch keineswegs trivial, denn häufig werden mehrere, möglicherweise miteinander konkurrierende Ziele gleichzeitig betroffen.

So mag eine Maßnahme beispielsweise bei den Betroffenen die Löhne steigern, sofern sie beschäftigt bleiben, jedoch die individuelle Beschäftigungswahrscheinlichkeit vermindern (man denke z.B. an die Einführung eines Mindestlohns im Niedriglohnsektor). Oder im Vergleich zweier Maßnahmen mag die eine (z.B. eine Qualifizierungsmaßnahme) die Löhne der Beschäftigten sehr stark anheben, während die andere (z.B. eine Lohnsubvention) eher auf die Beschäftigungswahrscheinlichkeit wirkt. In jedem dieser Fälle müssten demnach Ziele

[5] Interventionen, die auf einer aggregierten Ebene ansetzen, beispielsweise regionale Infrastrukturmaßnahmen, betreffen Individuen auf dieser Aggregationsebene in gleicher Form und lassen daher keinen Vergleich von Betroffenen und Nicht-Betroffenen innerhalb von Regionen zu. In diesem Falle müssen dann entweder intraregionale Korrelationen in den Zielvariablen statistisch berücksichtigt werden (siehe Schmidt et al. (1999) bzw. Augurzky und Schmidt (2000)) oder die Analyse wird sogleich auf der Ebene der Regionen durchgeführt (für das Beispiel der Reform der Arbeitsförderung im Rahmen des SGB III siehe Fertig und Schmidt (2000)).

gegeneinander abgewogen werden. In Zeiten hoher Arbeitslosigkeit scheint die Antwort vordergründig nahezuliegen, aber eine lexikographische Präferenz für eine Senkung der Arbeitslosigkeitsrate kann eigentlich kaum sinnvoll sein. Es verbleibt also grundsätzlich ein gesellschaftliches Entscheidungsproblem jenseits der Evaluation -- eine Aufgabe, die am besten durch die Berücksichtigung aller relevanten Zielgrößen vorbereitet wird.

Sollte sich anhand des Vergleichs von tatsächlicher und kontrafaktischer Situation ein merklicher Effekt der Maßnahme feststellen lassen, so ist ein weiterer Vergleich erforderlich, der diesen Effekt mit den entstehenden Kosten, direkten wie auch indirekten, konfrontiert. Diese Frage nach der Kosteneffektivität wirft ebenfalls methodische Probleme auf, da die direkt entstehenden Kosten, z.B. die gezahlten Lohnsubventionen, nur einen Teil der Kosten umfassen. Je nach Maßnahme fallen darüber hinaus Opportunitätskosten bei den Arbeitnehmern und Betrieben an -- so kann ein Teilnehmer einer Fortbildungsmaßnahme in dieser Zeit keiner anderen Tätigkeit nachgehen, einem Betrieb mögen durch die verwaltungstechnischen Aspekte der Lohnkostensubventionierung zusätzliche Verwaltungskosten entstehen. Darüber hinaus ist der staatliche Verwaltungsapparat ja keineswegs kostenlos, was uns die augenblickliche Diskussion des Umfangs staatlicher Aktivitäten eindeutig in Erinnerung ruft. Eine Vernachlässigung gerade der Kosten der Implementierung arbeitsmarktpolitischer Programme würde daher einen Weg in die wirtschaftspolitische Fehlsteuerung eröffnen.

In einer modernen Wirtschaft, mit ihren stetig wachsenden realen Produktionsmöglichkeiten und ihren schwer zu prognostizierenden und zu erkennenden konjunkturellen Schwankungen um diesen Wachstumspfad, aber auch mit ihrer heterogenen Struktur über Regionen und Sektoren hinweg, ist es häufig äußerst schwer anzugeben, wann eine Zielgröße in der Tat einen „vergleichsweise hohen“ Wert realisiert und wann nicht. Ein wohlbekanntes Phänomen, das diesen Punkt zu illustrieren vermag, ist die saisonale Entwicklung der Arbeitslosigkeitsrate im Jahresverlauf. Findet man bei einer arbeitsmarktpolitischen Maßnahme, beispielsweise einem Lohnsubventionierungsprogramm für Langzeitarbeitslose, die im Frühjahr eingesetzt wird, unter den Betroffenen eine auf den ersten Blick ansprechende Zahl von Arbeitnehmern, die eine neue Arbeitsstelle antreten, so ist grundsätzlich zu fragen, ob es nicht der übliche saisonale Aufschwung gewesen ist, der zumindest einem Teil von ihnen diesen Erfolg ermöglichte. Im Gegensatz dazu mag ein zahlenmäßig relativ moderater Effekt einer solchen Maßnahme zum Jahresende hin auf den ersten Blick desillusionierend wirken, aber unter Berücksichtigung des saisonalen Musters der Abgangsraten aus der Arbeitslosigkeit in der Tat einen starken Effekt des Eingriffs andeuten. Es ist also die Konstruktion einer angemessenen Vergleichssituation, mit der eine überzeugende Evaluierung der Maßnahme steht und fällt.

Um bei dieser Konfrontation mit der tatsächlichen Situation den Effekt der Maßnahme überzeugend abschätzen zu können, sollte sich die zur Analyse herangezogene Vergleichssituation idealerweise lediglich in genau einem Umstand von der tatsächlich beobachteten Situation unterscheiden, dem, daß die Maßnahme nicht durchgeführt wurde. Im angeführten Beispiel liegt nahe, daß, wollte man mit der Wirksamkeit eines arbeitsmarktpolitischen Programms glänzen, man dieses Programm dann wohl in Zeiten konjunkturellen bzw. saisonalen Aufschwungs durchführen sollte. Der vermeintliche Erfolg wäre dann auch scheinbar anhand aller buchhalterischen Größen wie beispielsweise der Wiedereingliederungsquote der Betroffenen eindeutig zu belegen. Eine erfolgreiche Evaluierung des Eingriffs würde eine solche Motivation jedoch aufdecken, indem der angemessene Vergleich mit einer Situation gesucht wird, die die gleiche saisonale Komponente bzw. den gleichen Stand des Konjunkturzyklus widerspiegelt.

Es ist offensichtlich, daß die Konstruktion einer solchen Vergleichssituation typischerweise ein komplexes Unterfangen darstellt, müssen doch viele Einflußgrößen gleichzeitig Berücksichtigung finden. Nicht jedoch der Umstand, daß Vergleichssituationen durch eine große Zahl von zu berücksichtigenden Aspekten gekennzeichnet sind -- so sehr sich dadurch auch praktische Probleme ergeben -- ist die zentrale Erkenntnis, der sich die Evaluierung von arbeitsmarktpolitischen Eingriffen stellen muß, sondern die Anforderung, daß diese kontrafaktische Situation *konstruiert* werden muß. Sie kann, wie im nächsten Unterabschnitt formal dargestellt, schlichtweg nicht beobachtet werden und kann somit auch nicht durch rein betriebswirtschaftliche Erhebungstechniken nachgestellt werden. Dieses Phänomen wird daher in der Literatur auch als das *fundamentale Evaluierungsproblem* bezeichnet.

Die Evaluierung der Effekte staatlicher Maßnahmen ist demnach ein genuin empirisches Problem, das somit die Wahl einer geeigneten empirischen Strategie unabdingbar macht. Glücklicherweise steht -- wenn man willens ist, sich von der Illusion zu befreien, man könne allein mit verwaltungstechnischen Erhebungsmethoden sinnvolle Evaluierung betreiben -- ein umfangreiches wissenschaftliches Instrumentarium zur Verfügung, das zumindest grundlegende Richtlinien für ein überzeugendes Vorgehen in einer solchen Situation vorgibt. Der zentrale Eckpfeiler empirischen wissenschaftlichen Arbeitens neben den Anforderungen, alle Datenquellen eindeutig und sorgfältig offenzulegen, alle möglicherweise verbleibenden Fehlerquellen zu diskutieren und alle Daten und Ergebnisse öffentlich zugänglich zu machen, ist die dominante Rolle, die der empirischen Evidenz zugewiesen wird: da die „richtige“ Antwort auf die Frage nach dem Effekt eines Eingriffs nie gefunden werden kann, bedarf es des Gewichts der qualitativ (also in termini der Richtung und der ungefähren Größenordnung des Effekts, nicht jedoch im Hinblick auf die x-te Stelle nach dem Komma) übereinstimmenden Resultate einer Reihe von unabhängigen

Wissenschaftlern separat durchgeführter empirischer Studien, um davon sprechen zu können, man habe eine zuverlässige Antwort anzubieten.

Die Wahl einer geeigneten empirischen Strategie ist somit grundsätzlich der zentrale methodische Schritt, nachdem die relevanten Beobachtungseinheiten, Zielgrößen und Kostenelemente festgelegt wurden. In der modernen Ökonometrie und Statistik hat sich bei diesem Gestaltungs- und Entscheidungsproblem eine Vorgehensweise entwickelt, die strikt zwischen den hypothetischen Möglichkeiten eines Ansatzes und seiner Leistungsfähigkeit in der empirischen Praxis trennt. So wird in der Regel eine mögliche Vorgehensweise nur dann als geeignete Kandidatin angesehen, wenn sie zumindest unter hypothetischen idealen Bedingungen -- insbesondere einer unendlichen Stichprobengröße -- die richtige Antwort liefern würde. Dann spricht man in technischen Termini davon, daß diese Strategie dazu in der Lage ist, die gesuchte Größe zu *identifizieren*. Die gewählte Strategie wird daher auch häufig als *Identifikationsstrategie* bezeichnet. Die zentrale Rolle dieser Wahl ins Bewußtsein empirisch arbeitender Forscher zu rufen, ist der primäre Beitrag der modernen Evaluierungsliteratur in der Statistik und der Ökonometrie. Wie sich im folgenden zeigen wird, ist es häufig nur das randomisierte Experiment, das diesen harten Anforderungen gerecht wird -- aber auch nicht-experimentelles Vorgehen mag bisweilen konzeptionell überzeugend sein.

In der Praxis sind ideale Studienbedingungen natürlich nie gegeben. Auch wenn aufgrund einer geschickt gewählten Identifikationsstrategie zu erwarten ist, daß die in der Analyse ermittelte Antwort lediglich unsystematisch um den korrekten Wert schwankt, so verbleibt allein schon aufgrund der Endlichkeit der untersuchten Stichprobe ein gewisses Zufallselement. Die Abschätzung seiner Größenordnung ist eine weitere Herausforderung an die empirische Arbeit -- sie kann jedoch nie losgelöst von der Identifikationsstrategie gesehen werden und ist dieser Entscheidung nachgeordnet. Aufgrund der zentralen Rolle der Konstruktion des Kontrafaktischen wird diesen Punkten in den folgenden Unterabschnitten eine erhöhte Aufmerksamkeit geschenkt. Dabei werden der formale Rahmen und die Umsetzung dieses methodischen Gerüsts in der Evaluierungspraxis getrennt berücksichtigt.

3.2 Die kontrafaktische Situation

In diesem Unterabschnitt wird in kompakter Form der formale Rahmen eingeführt, der sich in jüngster Zeit in der statistischen und ökonometrischen Literatur zur Evaluierung staatlicher Eingriffe als einheitliches Denkgerüst herausgebildet hat. Die relevanten Untersuchungseinheiten (im folgenden konzentrieren sich der Einfachheit halber alle Beispiele auf Arbeitnehmer, die Übertragung auf Betriebe odere andere Untersuchungseinheiten ist jedoch problemlos) werden

dabei auf ihre zentralen Charakteristika reduziert; der Zustand, von einer Maßnahme betroffen zu sein wird mit „1“, der, von ihr unberührt zu bleiben, mit „0“ angezeigt, wobei die entsprechende Indikatorvariable mit D_i bezeichnet wird. Somit wird für einen Arbeitnehmer i , der Zugang zu einer Lohnsubvention geniesst, dieser Umstand durch $D_i = 1$ angezeigt. Zur Bewertung der Wirkung dieses Programms würde man nun für dieses Individuum i gerne einen Vergleich zwischen dem relevanten beobachteten Arbeitsmarktergebnis, z.B. Beschäftigungserfolg ein Jahr nach der Maßnahme ($D_i = 1$), mit dem hypothetischen Arbeitsmarkterfolg, der sich ohne dieses Programm (also für $D_i = 0$) ergeben hätte, durchführen.

Die relevante Zielgröße nach dem Eingriff soll hier als Y_i+D_i bezeichnet werden, falls Arbeitnehmer i Teilnehmer am Programm ist (z.B. da der Arbeitnehmer in einer Region lebt, in der das Programm implementiert wird), und als Y_i, wenn nicht. Der kausale Effekt der Maßnahme auf den Arbeitsmarkterfolg von Arbeitnehmer i wird somit durch die Größe D_i reflektiert[6]. Was im vorangegangenen Unterabschnitt als das *fundamentale Evaluierungsproblem* bezeichnet wurde, läßt sich nun auch formal charakterisieren: Es ist schlichtweg unmöglich, für einen gegebenen Arbeitnehmer sowohl Y_i als auch $Y_i + D_i$ gleichzeitig zu beobachten, sondern lediglich eines der beiden Ergebnisse, da ein und derselbe Arbeitnehmer nur entweder an einem Arbeitsmarktprogramm teilnehmen kann, oder nicht. Die Zielgröße, die sich ohne das Angebot der Förderung des Arbeitnehmers durch das entsprechende Arbeitsamt ergeben hätte, Y_i, ist somit das *kontrafaktische* -- und folglich nicht beobachtbare -- Ergebnis für Individuen im Programm, also für solche, für die $D_i = 1$ gilt (und umgekehrt).

Somit kann man nicht hoffen, daß man jemals in der Lage sein wird, individuenspezifische Wirkungen einer arbeitsmarktpolitischen Maßnahme auch nur annähernd zu ermitteln. Daher wird sich, im völligen Bewußtsein der möglichen Schwankungen der individuellen Effekte zwischen den geförderten Arbeitnehmern, das Interesse jeder Evaluierungsstudie auf zusammenfassende Größen, sog. *Evaluationsparameter*, konzentrieren. Diese unterscheiden sich vor allem dadurch, daß sie die Maßnahmeneffekte unterschiedlich zusammengesetzter Gruppen von Individuen zusammenfassen. Der prominenteste, jedoch keineswegs ausschließlich interessante Evaluationsparameter ist der Populationsdurchschnitt (die mathematische Erwartung $E(.)$) aller individuenspezifischen Effekte für solche Arbeitnehmer, die die Förderung auch tatsächlich erfahren haben, der sog. *mean effect of treatment on the treated*,

[6] Quasi im Vorübergehen wird hier implizit die wichtige -- und häufig kritische -- Annahme getroffen, daß die Teilnahme bzw. Nicht-Teilnahme anderer Arbeitnehmer an der Fördermaßnahme für den Effekt des Eingriffs auf den Arbeitsmarkterfolg des Individuums i unerheblich ist. In technischen Termini handelt es sich bei dieser Annahme um die *stable unit treatment value assumption* oder kurz um SUTVA (Rubin (1986)).

(1) $E(D|D=1) = E(Y+D|D=1) - E(Y|D=1)$.

Es handelt sich dabie um den Erwartungswert des individuenspezifischen Effekts der am Programm teilnehmenden Individuen.

Eine Schätzung des ersten Terms auf der rechten Seite obiger Gleichung, $E(Y+D|D=1)$, läßt sich ohne weiteres in einer gegebenen Stichprobensituation konstruieren, schließlich werden die Arbeitsmarktergebnisse der geförderten Arbeitnehmer ja beobachtet. Somit kann man darauf vertrauen, daß die Annäherung der stichprobenbasierten Schätzung an den wahren Wert bei steigender Stichprobengröße tendenziell immer genauer wird. Dahingegen kann ein Schätzer für den kontrafaktischen Erwartungswert $E(Y|D=1)$ nicht ohne weitere Annahmen gefunden werden, denn für keinen der Arbeitnehmer, die unter die Maßnahme fallen ($D_i=1$) werden die entsprechenden kontrafaktischen Arbeitsmarktergebnisse Y_i beobachtet. Somit muß ein Ersatz für $E(Y|D=1)$ gefunden werden, welcher den gleichen Wert wie der kontrafaktische Erwartungswert annimmt, jedoch aus beobachtbaren Stichprobenelementen (aus einer geeigneten Teilmenge der Beobachtungen von Nicht-Teilnehmern mit $D_i =0$) konstruiert werden kann.

Naturgemäß läßt sich eine solche Ersatzgröße nur unter Annahmen finden, die über die Informationen des bisher skizzierten formalen Rahmens hinausgehen. So würde beispielsweise die (sehr starke) Annahme $E(Y|D=1)= E(Y|D=0)$ die Schätzung der gesuchten Größe aus einer einfachen Stichprobe von Nicht-Teilnehmern ($D_i =0$) erlauben. Ebenso wie für $E(Y+D|D=1)$ bei den Teilnehmern wird die Annäherung der Schätzungen an den wahren Wert dann mit steigender Stichprobengröße tendenziell immer genauer ausfallen. Diese Annahme erlaubt somit die Konstruktion des Kontrafaktischen, sie ist eine sog. *Identifikationsannahme*, sie folgt jedoch keineswegs direkt aus den obigen formalen Ausführungen -- dies gilt ebenso wenig für alle denkbaren alternativen Identifikationsannahmen. Darüber hinaus läßt sie sich nicht statistisch testen: ein Vergleich der Mittelwerte der Y_i zwischen Teilnehmern und Nicht-Teilnehmern an der Maßnahme läßt sich nicht durchführen, denn für Teilnehmer werden die entsprechenden Werte ja gerade nicht beobachtet (sonst gäbe es das Evaluationsproblem erst gar nicht).

Ein wichtiges Element bei der Suche nach geeigneten Identifikationsannahmen ist die statistische Kontrolle für beobachtbare Heterogenität, d.h. die Berücksichtigung solcher beobachtbarer Einflußgrößen, die die Untersuchungseinheiten *vor* Beginn der Maßnahme charakterisieren und sowohl einen Einfluß auf die Teilnahme an der Maßnahme als auch auf deren Effekt ausüben. Zu diesen Einflußgrößen gehören beispielsweise demographische Größen wie Alter und Geschlecht und arbeitsmarktrelevante persönliche Eigenschaften wie der Ausbildungsstand, aber auch der Arbeitsmarkterfolg vor Beginn der Maßnahme. Auch wenn zwischen den Stichproben von Teilnehmern und Nicht-Teilnehmern keine direkte Vergleichbarkeit

herrscht (beispielsweise mögen die Teilnehmer im Durchschnitt schlechter ausgebildet sein), läßt sich dann durch eine geeignete Gewichtung der Stichprobe der Nicht-Teilnehmer eine solche Vergleichbarkeit herstellen.

Typischerweise unbeobachtet bleiben Eigenschaften wie Fleiß, Motivation oder Durchhaltevermögen. Üben diese Größen jedoch einen fühlbaren Einfluß sowohl auf die Maßnahmenteilnahme als auch auf die Arbeitsmarktergebnisse aus, so führt dies zum sog. *Selektionsproblem* -- die statistische Kontrolle für beobachtbare Variablen ist dann nicht mehr hinreichend. Die überzeugende Behandlung dieses Problems ist das Kernstück der ökonometrischen und statistischen Forschungsbemühungen im Bereich der Evaluationsliteratur.

Wie im nächsten Unterabschnitt diskutiert werden wird, ist es eine zentrale Aufgabe bei der Suche nach einer geeigneten Evaluierungsstrategie, weitere Informationen heranzuziehen, die die Rechtfertigung für die benötigten zusätzlichen Identifikationsannahmen liefern. Es kann sich bei diesen Informationen um solche zum Design der Studie handeln -- experimentelle Evaluationsstudien sind beispielsweise häufig konzeptionell überzeugende Ansätze --, oder um Erkenntnisse zum Verhalten der Arbeitnehmer, z.B. im Hinblick auf die Determinanten der Neigung, an der Maßnahme teilzunehmen oder nicht. Was jedoch in jeder endlichen Stichprobensituation verbleiben wird, sind *unsystematische* Schwankungen um den tatsächlichen Evaluationsparameter, eine Tatsache, mit deren Existenz wissenschaftliches Arbeiten leben muß.

3.3 Empirische Umsetzung

Wie bereits im vergangenen Unterabschnitt angedeutet wurde, lassen sich bereits in der Phase des Studiendesigns zwei grundlegende Ansätze unterscheiden. Die experimentelle Analyse stellt in den Naturwissenschaften den dominanten Ansatz dar und erfreut sich auch im sozialwissenschaftlichen Kontext breiter Zustimmung. Nicht zuletzt die jüngst durchgeführte Evaluierung einer der umfassendsten arbeitsmarktpolitischen Maßnahmen der USA, des JTPA-Programms, bezeugt ihre Bedeutung auch in der internationalen Evaluierungspraxis. Ihre konzeptionelle Stärke bezieht die experimentelle Analyse aus der Tatsache, daß sie -- unter der Bedingung, daß das Experiment ein genaues Abbild der tatsächlichen Implementierung einer Maßnahme „im Feld“ darstellt – einen überzeugenden Ansatz zur Bewältigung des geschilderten Identifikationsproblems darstellt.

Kernstück jedes Experiments ist die zufallsgesteuerte Auswahl von Untersuchungseinheiten (hier weiterhin Arbeitnehmer) in eine Teilstichprobe von Teilnehmern an der Maßnahme, die sog. *treatment group*, und eine Teilstichprobe solcher Arbeitnehmer, die gerne an der Maßnahme teilgenommen hätten, denen jedoch der Zugang aufgrund der Zufallsauswahl verwehrt bleibt. Diese sog. *control group* unterscheidet sich also -- wie bei der Charakterisierung der idealen

kontrafaktischen Vergleichssituation gewünscht -- lediglich in der Teilnahme an der Maßnahme, aber in keiner anderen Komponente, beobachtbar wie unbeobachtbar, systematisch von den Mitgliedern der treatment group. Für die konzeptionelle Güte dieses Ansatzes spielt in diesem Fall nicht die geringste Rolle, welche Eigenschaften die Teilnehmer an der Maßnahme aufweisen, ob sie beispielsweise gut oder schlecht ausgebildet sind. Aufgrund der Randomisierung ergibt sich, zumindest in hinreichend dimensionierten Stichproben und bei Aufrechterhaltung der kontrollierten Studienbedingungen, eine Balance aller relevanter Größen zwischen den beiden Teilstichproben. Die Berechnung der geschätzten Effekte ist dann alles andere als eine komplizierte Übung, simple Durchschnittsbildung reicht in der Regel aus[7].

Trotz ihres konzeptionell überlegenen Designs können sozialwissenschaftliche Experimente häufig nicht durchgeführt werden[8]; auch in anderen Anwendungsgebieten wie beispielsweise der Medizin ist diese Möglichkeit vielfältig nicht gegeben (siehe z.B. Schmidt et al. (1999)). Die Gründe liegen im politischen oder ethischen Bereich oder sind profanerer Natur. So ist die Implementierung von Experimenten häufig aus logistischen oder finanziellen Gründen problematisch, oder das zu evaluierende Programm ist bereits angelaufen, bevor man an die Durchführung der Evaluierung zu denken beginnt. Um ein sozialwissenschaftliches Experiment durchzuführen, bedarf es zudem eines entsprechenden Anbieters -- in den USA haben sich mehrere privatwirtschaftliche Anbieter etabliert, deren Expertise in Planung, Durchführung und Auswertung von sozialwissenschaftlichen Experimenten mittlerweile erheblich ist. Ein entsprechender Erfahrungsschatz ist in der Bundesrepublik naturgemäß nicht vorhanden.

In diesen Fällen bleibt nichts anderes, als eine nicht-experimentelle Evaluierungsstrategie zu suchen. Geschickterweise sollte diese Strategie so beschaffen sein (siehe z.B. Rosenbaum (1995)), daß die konzeptionellen Eigenschaften eines Experiments -- die Balance aller relevanten Einflußgrößen zwischen den Stichproben der Teilnehmer und der Nicht-Teilnehmer -- so gut wie möglich nachgeahmt werden, wiederum eine Frage der geeigneten Identifikationsannahmen. Unter dem Eindruck der vorliegenden Evidenz läßt sich zum gegenwärtigen Zeitpunkt das folgende vorläufige Fazit zur jeweiligen Rolle experimenteller und nicht-experimenteller Evaluierungsforschung ziehen:

Nicht-experimentelle Evaluierungsstudien der wichtigsten nordamerikanischen Ausbildungsprogramme zeichnen sich durch eine hohe Variabilität der Ergebnisse aus. Dies liegt in überwiegendem Maße an der Variabilität der Identifikationsannahmen in an sich vergleichbaren Analysesituationen. Insbesondere die geeignete Auswahl der zu berücksichtigenden Merkmale, die

[7] Das bedeutet allerdings keinerlei Abstrich an der Wissenschaftlichkeit des Vorgehens.

die beteiligten Individuen -- Teilnehmer und Nicht-Teilnehmer an der Maßnahme -- in der Zeit vor dem Programm auszeichneten, kann die Forschungsergebnisse erheblich beeinflussen. Werden hier entscheidende Fehler begangen, beispielsweise durch die ungeeignete Berücksichtigung zeitlich verknüpfter Ergebnisvariablen, so mögen für den Forscher unbeobachtbare individuelle Charakteristika zu einer starken Verzerrung der Ergebnisse führen.

In jüngerer Zeit wurden jedoch einige fortgeschrittene nicht-experimentelle Ansätze entwickelt, die in ihrer Leistungsfähigkeit durchaus an die konzeptionell überlegenen sozialwissenschaftlichen Experimente heranreichen. Eine theoretische Diskussion dieses Abwägungsproblems bieten Schmidt et al. (1999), wohingegen Augurzky und Schmidt (2000) diese Abwägung in unterschiedlichen Evaluierungssituationen diskutieren und numerische Illustrationen anbieten. Nichtsdestoweniger darf man getrost bezweifeln, daß solche nicht-experimentellen Ansätze in näherer Zukunft experimentelle Ansätze konzeptionell vollkommen ablösen können. Im Gegenteil, obwohl nicht-experimentelle Techniken immer eine bedeutende Rolle spielen werden, beispielsweise bei einer Extrapolation bestehender Ergebnisse aus dem Erfahrungsschatz eines spezifischen Programms heraus in eine Situation, die durch andere Zielgruppen, andere Maßnahmenbündel oder unterschiedliche konjunkturelle Zustände charakterisiert ist, so bleiben sozialwissenschaftliche Experimente dennoch der verlässlichste und konzeptionell überzeugendste Ansatz zur Programmevaluierung (siehe auch Schmidt (2000b)).

Wie bereits oben angedeutet ist keine empirische Strategie in der Lage, die Unsicherheit über die korrekte Antwort völlig aufzulösen. Stattdessen verbleibt in jeder Schätzung eine Unsicherheit, deren Größenordnung in jeder soliden wissenschaftlichen Studie zusammen mit der Schätzung des Effektes vom Analytiker zu ermitteln ist. Dabei ist die verbleibende Unsicherheit auch ein Resultat der Schärfe der getroffenen Identifikationsannahmen. Je strikter diese Annahmen ausfallen, desto geringer wird tendenziell die verbleibende Unsicherheit ausgewiesen. Sind die Annahmen unzutreffend -- was, wie oben an einem Beispiel illustriert, zumindest für bindende Restriktionen nicht zu testen ist --, dann wird die verbleibende Unsicherheit häufig vermeintlich relativ gering sein.

Darüber, ob es sich allerdings um ein erfreulich präzises Resultat oder um eine verzerrte Antwort handelt, kann jedoch kein statistischer Test entscheiden, sondern nur ein kritisches Überdenken der Identifikationsstrategie. Es ist diese Abhängigkeit von den Identifikationsannahmen, die zumindest ein Grundverständnis der Möglichkeiten und Grenzen von

[8] Arbeitsmarktpolitische Modellversuche werden in der öffentlichen Diskussion ebenfalls häufig als „Experiment“ bezeichnet. In diesem Zusammenhang bezieht sich dieser Ausdruck lediglich auf die nicht-allumfassende und vorläufige Form des Programms, keineswegs auf das Studiendesign zu seiner Evaluierung.

Evaluierungsstrategien für jeden Auftraggeber in Politik und Verwaltung unabdingbar machen und eine kritische Distanz zu jeglichen allzu eindeutigen Ergebnissen nahelegen.

4 Bisherige empirische Evidenz

Die umfangreiche Erfahrung mit Arbeitsmarktprogrammen und ihrer Evaluierung, die in den USA in den letzten Jahrzehnten gesammelt wurde, ist der verläßlichste Ausgangspunkt für eine Sichtung der Evidenz zur Wirksamkeit arbeitsmarktpolitischer Maßnahmen (Überblicksartikel sind Friedlander et al. (1997), Heckman et al. (1999), Schmidt (2000c) und Stanley et al. (1998)). Allerdings gibt es auch in ganz Europa eine wachsende Zahl methodisch anspruchsvoller und dadurch im Vergleich mit ihren Vorgängern inhaltlich überzeugenderer Evaluationsstudien, im Einklang mit der Vielzahl innovativer Versuche, durch staatliche Interventionen die wirtschaftliche Situation von Teilnehmern dieser Maßnahmen zu verbessern (einen Überblick geben Kluve und Schmidt (2000), stellvertretend für die jüngere Forschung in der Bundesrepublik sei Lechner (1998) genannt). Zwar läßt sich dabei in mancher Hinsicht ein gewisser „europäischer Charakter“ erkennen, etwa in der weitgehenden Konzentration auf Arbeitsmarktprogramme für Jugendliche, die Betonung von Beschäftigungs- gegenüber Einkommenseffekten als Zielgrößen oder den nahezu völligen Ausschluß sozialwissenschaftlicher Experimente. Unglücklicherweise sieht es aber sehr danach aus, daß viele der früheren optimistischen Resultate ein zu positives Bild der Leistungsfähigkeit von Arbeitsmarktprogrammen in Europa gezeichnet hatten.

Insgesamt legt die vorhandene Evidenz nahe, daß solche Programme, seien es Ausbildungsmaßnahmen oder die Steuerung des Arbeitsmarktverhaltens durch finanzielle Anreize, bei Vorliegen einer Reihe günstiger Umstände durchaus in der Lage sind, ihr Ziel zu erreichen. Durch diese Maßnahmen geförderte Arbeitnehmer erzielen danach häufig ein höheres Einkommen und weisen höhere Einstellungschancen auf, ohne daß die Durchführung der Maßnahme prohibitiv hohe Kosten aufweisen muß. Diese positiven Effekte sind von der Größenordung her in der Regel recht bescheiden, halten aber häufig auch nach Ablauf des Programms eine Weile an. In jedem Falle zeichnet sich deutlich ab, daß es kein „Allheilmittel“ gibt, kein Arbeitsmarktprogramm, das für alle Zielgruppen wirtschaftlich benachteiligter Arbeitnehmer, über alle Regionen und Phasen des Konjunkturzyklus hinweg seine positive Wirkung zu entfalten in der Lage ist.

Unter den kosteneffektivsten Maßnahmen scheinen im nordamerikanischen Kontext vor allem die relativ kostengünstigen Vermittlungshilfen in den Arbeitsmarkt zu sein, aber auch die typischerweise weit kostenintensiveren Ausbildungsmaßnahmen und Maßnahmen zur Förderung der Arbeitsmarkterfahrung, die ihre spürbaren Kosten offenbar durch höhere Durchschnittseffekte einigermaßen ausgleichen können. Die tatsächliche Auswirkung jedes Programms hängt von der

konkreten Ausgestaltung der Maßnahme sowie vor allem von seiner jeweiligen Zielgruppe ab. Die am ehesten optimistisch stimmenden Resultate wurden übereinstimmend für benachteiligte erwachsene Frauen gefunden, sowohl für freiwillige als auch für verpflichtende Ausbildungsmaßnahmen im Rahmen von sog. *welfare-to-work* Programmen. Die Frage nach Programmauswirkungen auf benachteiligte erwachsene Männer hinterläßt größere Unsicherheit.

Eine mit der grundsätzlichen Heterogenität der Effekte einhergehende Beobachtung betrifft die relativ entmutigenden Resultate nahezu aller Ausbildungsprogramme, die auf die Verbesserung der Arbeitsmarktaussichten ökonomisch benachteiligter Jugendlicher abzielen. Es scheint offenbar kaum eine Möglichkeit zu geben, diesen zu helfen, wenn sie einmal die Entscheidung getroffen haben, ihre Schulausbildung nicht zu beenden. Will man daher die wirtschaftlichen Bedingungen dieser Zielgruppe erfolgreich verbessern, so gibt es offensichtlich nur einen Weg, den äußerst frühen Einstieg gefährdeter Jugendlicher in Programme mit intensiver perönlicher Betreuung und Beratung. Es ist allerdings äußerst problematisch, aus diesen Schlußfolgerungen zu den Wirkungen von staatlichen Programmen auf die ökonomischen Aussichten US-amerikanischer Jugendlicher allzu weitreichende Schlußfolgerungen für Jugendliche in anderen Ländern zu ziehen. Daher ist das Potential von Arbeitsmarktprogrammen zur Verbesserung der ökonomischen Situation benachteiligter Jugendlicher beispielsweise in Deutschland eine weitgehend offene Frage.

Eine weitere Einschränkung der positiven Schlußfolgerungen hinsichtlich der Leistungsfähigkeit vor allem von Ausbildungsmaß nahmen besteht darin, daß die beobachteten Einkommenszuwächse oft nicht auf gesteigerte Produktivität, also das eigentliche Ziel von Ausbildungsmaßnahmen, sondern zu einem großen Teil auf gesteigerte Arbeitsintensität nach Ablauf des Programms zurückzuführen sind. Darüber hinaus ist für den Erfolg von Arbeitsmarktmaßnahmen mit ausschlaggebend, welche Voraussetzungen die Programmteilnehmer vor allem im Hinblick auf ihre bereits vorhandene Ausbildung mitbringen. So zeichnet sich als stabiles Muster ab, daß gerade die mit einem Grundgerüst an Arbeitsmarktfertigkeiten ausgestatten Arbeitnehmer besonders von Ausbildungsmaßnahmen profitieren. Dagegen ist es sehr schwer, erfolgreiche Ausbildungsprogramme für diejenigen durchzuführen, die schlecht ausgebildet in die Maßnahme eintreten. Für Arbeitnehmer mit sehr geringen Startvoraussetzungen haben sich dahingegen Programme als besonders vielversprechend erwiesen, deren Inhalte lediglich auf den Einstieg in die Arbeitswelt abzielen.

Im Hinblick auf die Einrichtung finanzieller Anreize scheint eindeutig belegt, daß angemessen formulierte Programme in der Lage sind, bemerkenswert deutliche Verhaltensänderungen zu erzielen, die durchaus auch nach Ablauf des Programms langfristig wirken können. Die induzierte Verhaltensänderung wie auch die Verbesserung der

Einkommenssituation der Zielgruppe werfen dabei recht geringe Kosten auf, wobei der angemessene Vergleichsmaßstab die Ausgaben in einem Sozialversicherungssystem ohne Anreizprogramm sind. Besonders erfolgreich und kostengünstig scheinen solche Anreizprogramme zu sein, die Ausbildungskomponenten und die Ausprägung beruflicher Fertigkeiten mit Lohnsubventionen an Arbeitgeber kombinieren.

Auf der anderen Seite sind die Zielgruppen dieser Anreizschemata häufig durch solch ungünstige ökonomische Startbedingungen charakterisiert, daß auch ein erfolgreiches Programm ihre wirtschaftlichen Probleme nicht völlig zu lösen in der Lage ist. Ein zentraler Bestandteil erfolgreicher Anreizprogramme ist die Vermeidung von Mitnahmeeffekten, d.h. hoher Ausgaben an solche Individuen aus der Zielgruppe, die ihr Verhalten nicht aufgrund des Programms ändern, sog. *windfall beneficiaries*. Dies sind beispielsweise Arbeitnehmer, die eine Prämie dafür erhalten, die Abhängigkeit von der Sozialhilfe gegen eine Arbeitsstelle eingetauscht zu haben, obwohl sie dies auch ohne das Anreizprogramm getan hätten.

Natürlich ist es schwer, diese Individuen zu erkennen, daher müssen Kernaspekte des Programms so zugeschnitten sein, daß sie bloße Mitnahmeeffekte dieser Art möglichst vermeiden. Zu solchen Aspekten gehören längere Anwartschaftsperioden und Verpflichtungen zu einer Mindestzahl von Arbeitsstunden. Erfreulicherweise kann zum augenblicklichen Zeitpunkt bereits angedeutet werden, daß es durchaus gelingen kann, auch nach Ablauf der direkten Anreize durch Zuschußzahlungen die wirtschaftliche Situation der Zielgruppe dauerhaft zu verbessern. So wenig konkrete Evidenz auch insgesamt vorliegt, so scheint es doch nicht der Fall zu sein, das ein Auslaufen der Zuschüsse die Rückkehr ins soziale Netz bedeuten muß. Offenbar reicht häufig ein erster Anstoß aus, um danach der wirtschaftlichen Eigenverantwortung das Feld zu überlassen.

Ein wichtiges Element in der abschliessenden Bewertung jedes Anreizprogramms ist neben der Analyse der wirtschaftlichen Situation der Teilnehmer die Größenordnung der entstehenden Verdrängungseffekte. Somit ist die angemessene kontrafaktische Fragestellung hier etwas komplizierter, da sie ebenfalls berücksichtigen muß, wie Nicht-Teilnehmer ohne das Programm abgeschnitten hätten. Im allgemeinen läßt sich das Fazit ziehen, daß gerade diejenigen Programme erfolgreich sein dürften, die sich durch eine geringe administrative Komplexität auszeichnen.

Im Gegensatz zu diesen ermutigenden Schlußfolgerungen für Ausbildungsmaßnahmen und Anreizprogramme schneiden Arbeitsbeschaffungsmaßnahmen im Urteil der empirischen Evidenz eher schlecht ab. Zwar hat die Bereitstellung von zusätzlichen Arbeitsstellen durch die öffentliche Hand für klar abgegrenzte Gruppen von Arbeitnehmern durchaus ihren Platz als Übergangslösung in Zeiten außerordentlich großer wirtschaftlicher Probleme, aber es scheint in jedem Falle äußerst schwierig, sie so zu gestalten, daß eine umfassende Verschwendung von Ressourcen vermieden

wird. Die zentrale Frage in diesem Bereich ist: wird tatsächlich Arbeit geschaffen oder werden diese Bemühungen gänzlich durch Verdrängungseffekte zunichte gemacht?

Verdrängungseffekte nehmen dabei zwei grundsätzliche Formen an. Zum einen werden im öffentlichen Sektor Arbeitnehmer freigesetzt oder entgegen der ursprünglichen Planung nicht eingestellt, um für die Teilnehmer der Arbeitsbeschaffungsmaßnahme Platz zu machen (das Phänomen der sog. *public worker substitution*). Obwohl hier schon große Meßprobleme auftreten, so ist doch noch schwerer zu erfassen, wieviele Arbeitsplätze im privatwirtschaftlichen Bereich verloren gehen, weil die mit den günstigen Arbeitnehmern aus der Arbeitsbeschaffungsmaßnahme ausgestatteten Firmen die entsprechende Leistung günstiger anbieten können. Dies ist das Phänomen des sog. *private job crowd-out*.

Substitutionseffekte können durch geeignet konstruierte Maßnahmen recht gering gehalten werden, aber das gelingt offenbar nur in Programmen mit hohem administrativen Aufwand und harten Teilnahmebedingungen. Erfolgversprechende Programme sind typischerweise eng auf eine Zielgruppe zugeschnitten, sind grundsätzlich von zeitlich limitierter Natur und zahlen Niedriglöhne. Ähnliche Voraussetzungen müssen erfüllt sein, um Verdrängungseffekte im privatwirtschaftlichen Bereich weitgehend zu vermeiden. Insgesamt scheint die Schlußfolgerung nahe zu liegen, daß unter bestimmten Bedingungen auch Arbeitsbeschaffungsmaßnahmen eine legitime Rolle im Spektrum der Arbeitsmarktprogramme einnehmen können, insbesondere als Instrument temporärer Entlastung des Arbeitsmarkts, auch wenn viele Einschränkungen bestehen bleiben. Als dauerhafte Lösung oder gar als Instrument der kostengünstigen Bereitstellung gesellschaftlich wertvoller Güter und Dienstleistungen sind sie weitgehend ungeeignet.

Obwohl sie politisch so überaus relevant sind und teilweise bedeutende Anteile des öffentlichen Budgets für sich in Anspruch nehmen, wurden Arbeitsbeschaffungsmaßnahmen nur äußerst selten einer wissenschaftlichen Evaluierung unterzogen. Demnach besteht hier ein großes Forschungsdefizit. Wenn man bedenkt, daß die konzeptionellen Probleme dieser Programme von Beginn an recht offensichtlich gewesen sind, sie aber dennoch in vielen Volkswirtschaften eine lange Tradition aufweisen, dann ist es schon äußerst dramatisch, daß bislang so wenig „harte“ Erkenntnisse über sie gesammelt werden konnten. Es scheint jedoch eine recht gesicherte Position, davon auszugehen, daß Arbeitsbeschaffung im öffentlichen Sektor kein vielversprechender Ansatz zur Lösung der Arbeitsmarktprobleme wirtschaftlich benachteiligter Arbeitnehmer sein kann.

Schlußendlich sind die Rahmenbedingungen, die durch das ökonomische Umfeld gesetzt werden, vor allem aber die geltenden institutionellen Regelungen des Sozialversicherungssystems entscheidende Faktoren für die mögliche Wirksamkeit vor allem von Anreizprogrammen. So wird es beispielsweise äußerst schwer sein, durch ein Ausbildungsprogramm die zukünftige

wirtschaftliche Selbständigkeit von Teilnehmern zu fördern, wenn durch die Teilnahme an der Maßnahme eine ausgelaufene Anwartschaft auf Unterstützungsleistungen erneuert werden kann (siehe z.B. Góra und Schmidt (1998) und Kluve et al. (1999)). In diesem Fall mag der Hauptgrund für die Teilnahme die Rückkehr in die Gruppe der Leistungsberechtigten sein, und nicht die Akquisition weiteren Humankapitals.

5 Agenda

Nichts wäre den weniger erfolgreichen europäischen Arbeitnehmern mehr zu wünschen, als daß die politischen Entscheidungsträger ihre augenblicklichen, zarten Bestrebungen noch deutlich erhöhten, mehr über die Funktionsweise des Niedriglohnsektors, die Basis der Entscheidungen von Arbeitnehmern und Arbeitgebern in diesem Sektor und die Wirksamkeit unterschiedlichster Maßnahmen Aktiver Arbeitsmarktpolitik zu erfahren. Dabei sollte ein Anreiz für die Verantwortlichen in Politik und Verwaltung sein, daß die wissenschaftliche Begleitung innovativer Maßnahmen zur Verbesserung der wirtschaftlichen Situation europäischer Arbeitnehmer durchaus große Früchte tragen kann.

Andererseits ist auch eine gute Maßnahme ohne wissenschaftliche Evaluation nicht als solche erkennbar, da die Konstruktion der hypothetischen Vergleichssituation, die zur Bewertung der Maßnahme unabdingbar erforderlich ist, nicht durch bloße Prozeßanalyse, durch primitive Vorher-Nacher-Vergleiche oder konzeptionell ähnlich schwache Ansätze bewerkstelligt werden kann. In der Evaluationsforschung werden Experimente von vielen Spezialisten hauptsächlich deswegen favorisiert, weil sie einfache und gleichzeitig glaubwürdige Resultate liefern. Außerdem stehen in experimentellen Studien wichtige inhaltliche Themen im Mittelpunkt der Diskussion -- im Gegensatz dazu lenken nicht-experimentelle Studien die Aufmerksamkeit auf methodologische Fragen, wie etwa die Wahl der Vergleichsgruppe, die besonderen *Identifikationsannahmen*, die den Daten auferlegt werden, oder das spezifische Verhaltensmodell, das für die Erklärung der Programmteilnahme formuliert wurde. Sollen Entscheidungsträger und die breite Öffentlichkeit über die Auswirkungen staatlicher Interventionen verläßlich unterrichtet werden, so sind experimentelle Studien sicherlich die erste Wahl.

Zum augenblicklichen Zeitpunkt lassen sich einige klare Lektionen aus der bestehenden Evidenz ziehen, die bei der Suche nach geeigneten Vorschlägen für arbeitsmarktpolitische Eingriffe heranzuziehen sind. So sollte, wenn immer die Ausgestaltung und Umsetzung eines staatlichen Ausbildungsprogramms diskutiert wird, zuerst geklärt sein, welche Form des Marktversagens mit dieser Maßnahme korrigiert werden soll. Dabei kommt erschwerend hinzu, daß der Niedriglohnsektor bislang offenbar nicht in befriedigender Form modelliert werden konnte (siehe

z.B. die Kontroverse um die Wirkung gesetzlicher Mindestlöhne, die sich an Card und Krueger (1995) entzündet hat).

Obwohl sicherlich das Potential des Marktes nicht unterschätzt werden sollte, auch eine größere Ausweitung des Arbeitsangebots in Beschäftigung umzusetzen (denn nichts ist weiter von der Realität entfernt als die Vorstellung vieler Laien, es gäbe zu jedem Zeitpunkt nur eine bestimmte Anzahl von Arbeitsstellen, die dann irgendwie unter den mehr oder weniger Glücklichen aufgeteilt werden), so könnte es bei einem Nachlassen der konjunkturellen Stärke der Wirtschaft angebracht erscheinen, zuweilen zeitlich begrenzte Stellen im öffentlichen Sektor zu schaffen. Allerdings drohen hier nahezu unvermeidlich Verdrängungseffekte.

In jedem Falle steht anstelle eines punktuellen Eingriffs, der auf eine Zielgruppe benachteiligter Arbeitnehmer oder eine bestimmte Fehlsteuerung des Sozialversicherungssystems abzielt, grundsätzlich die Möglichkeit zur Verfügung, das System der sozialen Sicherung umzugestalten und flankierende Reformen zur Förderung der wirtschaftlichen Rahmenbedingungen durchzuführen. Dazu gehört natürlich ein höheres Vertrauen in den Markt, das offenbar in den angelsächsischen Ländern zu einem erheblichen Beschäftigungswachstum im Niedriglohnbereich geführt hat, wie auch eine Eindämmung des offensichtlich oft hemmenden verwaltungsjuristischen Einflusses auf den Arbeitsmarkt und auch auf die hier diskutierten arbeitsmarktpolitischen Eingriffe (auch eine solche Eindämmung wäre natürlich ein Evaluierungsproblem).

Bei jedem arbeitsmarktpolitischen Eingriff sollten praktische Umsetzbarkeit und möglichst geringe administrative Kosten besonders betont werden. Letztlich können auch geringe positive Effekte, die durch kostengünstige Maßnahmen hervorgerufen werden, hochgradig kosteneffizient sein, während andererseits großangelegte und teure Programme mit Ausbildungskursen erst beträchtliche Effekte hervorrufen müssen, um kosteneffizient zu sein. Es wäre auch im Kontext der Arbeitsbeschaffungsmaßnahmen fatal, die Kosten des administrativen Apparats aus den Überlegungen herauszulassen. Im Gegenteil, solche Verwaltungskosten sind offenbar ein substantieller Bestandteil jeglicher staatlicher Interventionen. Im Rahmen von Arbeitsmarktprogrammen droht ihre Bedeutung mehr oder weniger exponentiell mit der Anzahl der durch die verwaltenden Stellen zu treffenden Ermessensentscheidungen zu steigen. Da die Ausweitung sinnvoller Beschäftigungsmöglichkeiten sicherlich nicht der *komparative* Vorteil der öffentlichen Hand sein dürfte, scheint es grundsätzlich sehr schwer, ein Arbeitsbeschaffungsprogramm zu entwerfen, das einerseits geringe administrative Komplexität aufweist, aber andererseits so aufgebaut ist, daß nur geringe Verdrängungseffekte befürchtet werden müssen.

Es ist darüber hinaus sicherlich nicht realistisch, zu erwarten, daß der erzielbare Effekt durch ein staatliches Ausbildungsprogramm denjenigen privater Ausbildung deutlich übertreffen kann. Dies impliziert, daß der Effekt eines erfolgreichen staatlichen Ausbildungsprogramms angesichts der üblicherweise kurzfristig angelegten Programmstruktur nur moderat sein kann. In Abhängigkeit der Probleme, auf die das Programm gerichtet ist, kann es sein, daß der Effekt nicht ausreichend ist, um das Problem zu lösen. Wie auch bei Ausbildungsprogrammen kann man bei finanziellen Anreizschemata nur Auswirkungen in Größenordnungen erwarten, die der Höhe der gesetzten Anreize entsprechen. Es wäre recht naiv zu erwarten, daß eine bescheidende Lohnsubvention die Einstellungsentscheidungen von Arbeitgebern hinreichend beeinflussen kann, wenn deren Einstellungs- und Anlernkosten erheblich sind, oder wenn die Hauptursache für die geringe Einstellungsneigung die Sorge ist, in Zeiten wirtschaftlichen Abschwungs die eingestellten Arbeitnehmer nicht wieder freisetzen zu können. In solchen Situationen muß eine Lohnsubvention entsprechend bemessen sein, um eine Verhaltensänderung seitens des Arbeitgebers zu erzielen.

Eine ähnliche Überlegung gilt für Lohnsubventionen, die – was im Prinzip eher begrüßenswert ist -- an die Einstellung von Arbeitnehmern aus einer bestimmten Zielgruppe gebunden sind. Zieht die Identifikation als Teil dieser Zielgruppe ein soziales Stigma nach sich, so werden viele potentiell geförderte Arbeitnehmer lieber darauf verzichten, sich zu erkennen zu geben -- es sei denn, die Lohnsubvention ist erheblich. Schließlich mögen die ökonomischen Defizite der Zielgruppe auch derart substantiell sein, daß auch eine Ausweitung ihrer Beschäftigungsrate diese nicht völlig beheben kann.

Die Effekte staatlicher Programme sind typischerweise heterogen in bezug auf die Bevölkerung. Demnach kann eine sorgfältige Auswahl der Zielgruppe bereits vorbestimmen, was das Programm leisten kann. Die implizite Mißachtung dieser Erkenntnis war eine der Hauptursachen für das relativ enttäuschende Abschneiden größerer sozialwissenschaftlicher Experimente der 70er Jahre im Zeichen der sog. *Negative Income Tax*. Diese Experimente wurden als allgemeingültige Programme des sozialen Sicherungssystems in Zeiten eines expandierenden Wohlfahrtsstaats konzipiert. Im Gegensatz dazu betonen die innovativsten finanziellen Anreizprogramme der Gegenwart die Ermutigung zur Selbsthilfe und der wirtschaftlichen Eigenverantwortung der Zielgruppe.

5.1 Die Ausbildung in den Wirtschaftswissenschaften

Studenten der Volkswirtschaftslehre wie auch benachbarter Sozialwissenschaften geniessen als einen wichtigen Bestandteil ihres Grundstudiums üblicherweise eine umfassende Einführung in die Methoden der beschreibenden und der schliessenden Statistik. Unglücklicherweise ist es den

meisten Studierenden dieses Fachs an deutschen Fakultäten -- ganz im Gegensatz zu ihren Studienkollegen in den USA, dem Vereinigten Königreich oder Frankreich -- jedoch immer noch möglich, in ihrem weiterführenden Studium das zentrale Fach Ökonometrie gänzlich zu vermeiden. Dies ist in doppelter Hinsicht fatal. Die Arbeit einer Ökonomin ist nunmal weitgehend quantitativer Natur, unabhängig davon, ob sie in der volkswirtschaftlichen Abteilung einer Bank, in einer internationalen Organisation oder in einem Ministerium ihre berufliche Bestimmung findet. Empirische Techniken sind daher grundlegendes Handwerkszeug, sie nicht zu beherrschen in unserer globalisierten Wirtschaft und Gesellschaft ein wichtiger Wettbewerbsnachteil.

Darüber hinaus stehen in der angesprochenen Statistikausbildung naturgemäß die methodischen Grundlagen des Schätzens und Testens im Vordergrund. Die Anwendung dieser Techniken auf konkrete ökonomische Fragestellungen, insbesondere die Klärung der (nicht nur) im Bereich der Evaluierungsforschung so zentralen Identifikationsproblematik, ist nicht ihr Gegenstand. Daher droht, trotz der angesprochenen dramatischen Fortschritte der Erkenntnisse im Bereich der Evaluierung wirtschaftspolitischer Maßnahmen, auch die heutige Generation des wirtschaftswissenschaftlichen Nachwuchses die Gelegenheit zu verpassen, die Notwendigkeit zur statistisch-ökonometrischen Evaluierung dieser Maßnahmen und die Möglichkeiten und Grenzen des vorhandenen wissenschaftlichen Instrumentariums zu verstehen. Um diesem Dilemma abzuhelfen, muß sich die Ausbildung in den Wirtschaftswissenschaften weiter gegenüber empirischen Inhalten öffnen -- es gilt, den Nachwuchs anzuleiten, die Ökonometrie einerseits zu „entzaubern“ und mit ihr andererseits als einem natürlichen Verbündeten umzugehen. Die Qualität zukünftiger Wirtschaftspolitik kann davon nur profitieren.

Literatur

Augurzky, Boris und Christoph M. *Schmidt* (2000) The Evaluation of Community-Based Interventions. A Monte Carlo Study, mimeo. Universität Heidelberg.

Card, David und Alan B. *Krueger* (1995) Myth and Measurement, Princeton: Princeton University Press.

Fertig, Michael und Christoph M. *Schmidt* (2000) Discretionary Measures of Active Labor Market Policy: The German Employment Promotion Reform in Perspective, IZA Discussion Paper Nr. 182.

Friedlander, Daniel, David H. *Greenberg* und Philip K. *Robins* (1997) Evaluating Government Training Programs for the Economically Disadvantaged, Journal of Economic Literature 35, 1809-1855.

Góra, Marek und Christoph M. *Schmidt* (1998) Long-Term Unemployment, Unemployment Benefits and Social Assistance: The Polish Experience, Empirical Economics 23, 55-85.

Heckman, James J., Robert J. *LaLonde* und Jeffrey A. *Smith* (1999) The Economics and Econometrics of Active Labor Market Programs, in: *Ashenfelter*, Orley und David *Card* (Hrsg.) Handbook of Labor Economics vol. III, Amsterdam et al.: North-Holland, 1865-2097.

Kluve, Jochen, Hartmut *Lehmann* und Christoph M. *Schmidt* (1999) Active Labor Market Policies: Human Capital Enhancement, Stigmatization, or Benefit Churning?, Journal of Comparative Economics 27, 61-89.

Kluve, Jochen und Christoph M. *Schmidt* (2000)} Can Training and Financial Incentives Combat European Unemployment? A Survey of Recent Evaluation Studies, mimeo. Universität Heidelberg.

Lechner, Michael (1998) Training the East German Labor Force -- Microeconometric Evaluations of Continous Vocational Training After Unification. Heidelberg: Physica-Verlag.

Manski, Charles F. (1995) Identification Problems in the Social Sciences, Cambridge, Mass. et al.: Harvard Unversity Press.

Rosenbaum, Paul R. (1995) Observational Studies, New York: Springer Series in Statistics.

Rubin, Donald B. (1974) Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, Journal of Educational Psychology 66, 688-701.

Rubin, Donald B. (1986) Which Ifs Have Causal Answers?, Journal of the American Statistical Association 81, 961-962.

Schmidt, Christoph M. (1999) Knowing What Works. The Case for Rigorous Program Evaluation, IZA Discussion Paper Nr. 77.

Schmidt, Christoph M. (2000a) The Heterogeneity and Cyclical Sensitivity of Unemployment. An Exploration of German Labor Market Flows, ifo Studien 46, 73-98.

Schmidt, Christoph M. (2000b) Do We Need Social Experiments? The Potential and Limits of Non-experimental Project Evaluation, mimeo. Universität Heidelberg.

Schmidt, Christoph M. (2000c) Training, Incentives, and Public Service Jobs. The North American Experience With Labor Market Programs, mimeo. Universität Heidelberg.

Schmidt, Christoph M., Rob *Baltussen* und Rainer *Sauerborn* (1999) Evaluation of Community-Based Interventions: Group-Randomization, Limits and Alternatives, Dept. of Economics Discussion Paper No. 281, Universität Heidelberg.

Stanley, Marcus, Larry *Katz* und Alan B. *Krueger* (1998) Impacts of Employment and Training Programs: The American Experience, Background paper prepared for the British Chancellor of the Exchequer for the 1998 G-8 Meeting.

IZA Discussion Papers

No	Author(s)	Titel	Area	Date
111	V. Sorm K. Terrell	Sectoral Restructuring and Labor Mobility: A Comparative Look at the Czech Republic	1/4	2/00
112	L. Bellmann T. Schank	Innovations, Wages and Demand for Heterogeneous Labour: New Evidence from a Matched Employer-Employee Data-Set	5	2/00
113	R. Euwals	Do Mandatory Pensions Decrease Household Savings? Evidence for the Netherlands	3	2/00
114	G. Brunello A. Medio	An Explanation of International Differences in Education and Workplace Training	2	2/00
115	A. Cigno F. C. Rosati	Why do Indian Children Work, and is it Bad for Them?	3	2/00
116	C. Belzil	Unemployment Insurance and Subsequent Job Duration: Job Matching vs. Unobserved Heterogeneity	3	2/00
117	S. Bender A. Haas C. Klose	IAB Employment Subsample 1975-1995. Opportunities for Analysis Provided by the Anonymised Subsample	7	2/00
118	M. A. Shields M. E. Ward	Improving Nurse Retention in the British National Health Service: The Impact of Job Satisfaction on Intentions to Quit	5	2/00
119	A. Lindbeck D. J. Snower	The Division of Labor and the Market for Organizations	5	2/00
120	P. T. Pereira P. S. Martins	Does Education Reduce Wage Inequality? Quantile Regressions Evidence from Fifteen European Countries	5	2/00
121	J. C. van Ours	Do Active Labor Market Policies Help Unemployed Workers to Find and Keep Regular Jobs?	4/6	3/00
122	D. Munich J. Svejnar K. Terrell	Returns to Human Capital under the Communist Wage Grid and During the Transition to a Market Economy	4	3/00
123	J. Hunt	Why Do People Still Live in East Germany?	1	3/00
124	R. T. Riphahn	Rational Poverty or Poor Rationality? The Take-up of Social Assistance Benefits	3	3/00
125	F. Büchel J. R. Frick	The Income Portfolio of Immigrants in Germany - Effects of Ethnic Origin and Assimilation. Or: Who Gains from Income Re-Distribution?	1/3	3/00
126	J. Fersterer R. Winter-Ebmer	Smoking, Discount Rates, and Returns to Education	5	3/00

127	M. Karanassou D. J. Snower	Characteristics of Unemployment Dynamics: The Chain Reaction Approach	3	3/00
128	O. Ashenfelter D. Ashmore O. Deschênes	Do Unemployment Insurance Recipients Actively Seek Work? Evidence From Randomized Trials in Four U.S. States	6	3/00
129	B. R. Chiswick M. E. Hurst	The Employment, Unemployment and Unemployment Compensation Benefits of Immigrants	1/3	3/00
130	G. Brunello S. Comi C. Lucifora	The Returns to Education in Italy: A New Look at the Evidence	5	3/00
131	B. R. Chiswick	Are Immigrants Favorably Self-Selected? An Economic Analysis	1	3/00
132	R. A. Hart	Hours and Wages in the Depression: British Engineering, 1926-1938	7	3/00
133	D. N. F. Bell R. A. Hart O. Hübler W. Schwerdt	Paid and Unpaid Overtime Working in Germany and the UK	1	3/00
134	A. D. Kugler G. Saint-Paul	Hiring and Firing Costs, Adverse Selection and Long-term Unemployment	3	3/00
135	A. Barrett P. J. O'Connell	Is There a Wage Premium for Returning Irish Migrants?	1	3/00
136	M. Bräuninger M. Pannenberg	Unemployment and Productivity Growth**:** An Empirical Analysis within the Augmented Solow Model	3	3/00
137	J.-St. Pischke	Continuous Training in Germany	5	3/00
138	J. Zweimüller R. Winter-Ebmer	Firm-specific Training: Consequences for Job Mobility	1	3/00
139	R. A. Hart Y. Ma	Wages, Hours and Human Capital over the Life Cycle	1	3/00
140	G. Brunello S. Comi	Education and Earnings Growth: Evidence from 11 European Countries	2/5	4/00
141	R. Hujer M. Wellner	The Effects of Public Sector Sponsored Training on Individual Employment Performance in East Germany	6	4/00
142	J. J. Dolado F. Felgueroso J. F. Jimeno	Explaining Youth Labor Market Problems in Spain: Crowding-Out, Institutions, or Technology Shifts?	3	4/00
143	P. J. Luke M. E. Schaffer	Wage Determination in Russia: An Econometric Investigation	4	4/00
144	G. Saint-Paul	Flexibility vs. Rigidity: Does Spain have the worst of both Worlds?	1	4/00

145	M.-S. Yun	Decomposition Analysis for a Binary Choice Model	7	4/00
146	T. K. Bauer J. P. Haisken-DeNew	Employer Learning and the Returns to Schooling	5	4/00
147	M. Belot J. C. van Ours	Does the Recent Success of Some OECD Countries in Lowering their Unemployment Rates Lie in the Clever Design of their Labour Market Reforms?	3	4/00
148	L. Goerke	Employment Effects of Labour Taxation in an Efficiency Wage Model with Alternative Budget Constraints and Time Horizons	3	5/00
149	R. Lalive J. C. van Ours J. Zweimüller	The Impact of Active Labor Market Programs and Benefit Entitlement Rules on the Duration of Unemployment	3/6	5/00
150	J. DiNardo K. F. Hallock J.-St. Pischke	Unions and the Labor Market for Managers	7	5/00
151	M. Ward	Gender, Salary and Promotion in the Academic Profession	5	5/00
152	J. J. Dolado F. Felgueroso J. F. Jimeno	The Role of the Minimum Wage in the Welfare State: An Appraisal	3	5/00
153	A. S. Kalwij M. Gregory	Overtime Hours in Great Britain over the Period 1975-1999: A Panel Data Analysis	3	5/00
154	M. Gerfin M. Lechner	Microeconometric Evaluation of the Active Labour Market Policy in Switzerland	6	5/00
155	J. Hansen	The Duration of Immigrants' Unemployment Spells: Evidence from Sweden	1/3	5/00
156	C. Dustmann F. Fabbri	Language Proficiency and Labour Market Performance of Immigrants in the UK	1	5/00
157	P. Apps R. Rees	Household Production, Full Consumption and the Costs of Children	7	5/00
158	A. Björklund T. Eriksson M. Jäntti O. Raaum E. Österbacka	Brother Correlations in Earnings in Denmark, Finland, Norway and Sweden Compared to the United States	5	5/00
159	P.- J. Jost M. Kräkel	Preemptive Behavior in Sequential Tournaments	5	5/00
160	M. Lofstrom	A Comparison of the Human Capital and Signaling Models: The Case of the Self-Employed and the Increase in the Schooling Premium in the 1980's	5	6/00
161	V. Gimpelson D. Treisman G. Monusova	Public Employment and Redistributive Politics: Evidence from Russia’s Regions	4	6/00

162	C. Dustmann M. E. Rochina-Barrachina	Selection Correction in Panel Data Models: An Application to Labour Supply and Wages	6	6/00
163	R. A. Hart Y. Ma	Why do Firms Pay an Overtime Premium?	5	6/00
164	M. A. Shields S. Wheatley Price	Racial Harassment, Job Satisfaction and Intentions to Quit: Evidence from the British Nursing Profession	5	6/00
165	P. J. Pedersen	Immigration in a High Unemployment Economy: The Recent Danish Experience	1	6/00
166	Z. MacDonald M. A. Shields	The Impact of Alcohol Consumption on Occupational Attainment in England	5	6/00
167	A. Barrett J. FitzGerald B. Nolan	Earnings Inequality, Returns to Education and Immigration into Ireland	5	6/00
168	G. S. Epstein A. L. Hillman	Social Harmony at the Boundaries of the Welfare State: Immigrants and Social Transfers	3	6/00
169	R. Winkelmann	Immigration Policies and their Impact: The Case of New Zealand and Australia	1	7/00
170	T. K. Bauer K. F. Zimmermann	Immigration Policy in Integrated National Economies	1	7/00
171	C. Dustmann F. Windmeijer	Wages and the Demand for Health – A Life Cycle Analysis	5	7/00
172	D. Card	Reforming the Financial Incentives of the Welfare System	3	7/00
173	D. S. Hamermesh	Timing, Togetherness and Time Windfalls	5	7/00
174	E. Fehr J.-R. Tyran	Does Money Illusion Matter? An Experimental Approach	7	7/00
175	M. Lofstrom	Self-Employment and Earnings among High-Skilled Immigrants in the United States	1	7/00
176	O. Hübler W. Meyer	Industrial Relations and the Wage Differentials between Skilled and Unskilled Blue-Collar Workers within Establishments: An Empirical Analysis with Data of Manufacturing Firms	5	7/00
177	B. R. Chiswick G. Repetto	Immigrant Adjustment in Israel: Literacy and Fluency in Hebrew and Earnings	1	7/00
178	R. Euwals M. Ward	The Renumeration of British Academics	5	7/00
179	E. Wasmer P. Weil	The Macroeconomics of Labor and Credit Market Imperfections	2	8/00
180	T. K. Bauer I. N. Gang	Sibling Rivalry in Educational Attainment: The German Case	5	8/00

181	E. Wasmer Y. Zenou	Space, Search and Efficiency	2	8/00
182	M. Fertig C. M. Schmidt	Discretionary Measures of Active Labor Market Policy: The German Employment Promotion Reform in Perspective	6	8/00
183	M. Fertig C. M. Schmidt	Aggregate-Level Migration Studies as a Tool for Forecasting Future Migration Streams	1	8/00
184	M. Corak B. Gustafsson T. Österberg	Intergenerational Influences on the Receipt of Unemployment Insurance in Canada and Sweden	3	8/00
185	H. Bonin K. F. Zimmermann	The Post-Unification German Labor Market	4	8/00
186	C. Dustmann	Temporary Migration and Economic Assimilation	1	8/00
187	T. K. Bauer M. Lofstrom K. F. Zimmermann	Immigration Policy, Assimilation of Immigrants and Natives' Sentiments towards Immigrants: Evidence from 12 OECD-Countries	1	8/00
188	A. Kapteyn A. S. Kalwij A. Zaidi	The Myth of Worksharing	5	8/00
189	W. Arulampalam	Is Unemployment Really Scarring? Effects of Unemployment Experiences on Wages	3	8/00
190	C. Dustmann I. Preston	Racial and Economic Factors in Attitudes to Immigration	1	8/00
191	G. C. Giannelli C. Monfardini	Joint Decisions on Household Membership and Human Capital Accumulation of Youths: The role of expected earnings and local markets	5	8/00
192	G. Brunello	Absolute Risk Aversion and the Returns to Education	5	8/00
193	A. Kunze	The Determination of Wages and the Gender Wage Gap: A Survey	5	8/00
194	A. Newell F. Pastore	Regional Unemployment and Industrial Restructuring in Poland	4	8/00
195	F. Büchel A. Mertens	Overeducation, Undereducation, and the Theory of Career Mobility	5	9/00
196	J. S. Earle K. Z. Sabirianova	Equilibrium Wage Arrears: A Theoretical and Empirical Analysis of Institutional Lock-In	4	9/00
197	G. A. Pfann	Options to Quit	1	9/00

198	M. Kreyenfeld C. K. Spiess G. G. Wagner	A Forgotten Issue: Distributional Effects of Day Care Subsidies in Germany	3	9/00
199	H. Entorf	Rational Migration Policy Should Tolerate Non-Zero Illegal Migration Flows: Lessons from Modelling the Market for Illegal Migration	1	9/00
200	T. Bauer G. S. Epstein I. N. Gang	What are Migration Networks?	1	9/00
201	T. J. Dohmen G. A. Pfann	Worker Separations in a Nonstationary Corporate Environment	1	9/00
202	P. Francois J. C. van Ours	Gender Wage Differentials in a Competitive Labor Market: The Household Interaction Effect	5	9/00
203	J. M. Abowd F. Kramarz D. N. Margolis T. Philippon	The Tail of Two Countries: Minimum Wages and Employment in France and the United States	5	9/00
204	G. S. Epstein	Labor Market Interactions Between Legal and Illegal Immigrants	1	10/00
205	A. L. Booth M. Francesconi J. Frank	Temporary Jobs: Stepping Stones or Dead Ends?	1	10/00
206	C. M. Schmidt R. Baltussen R. Sauerborn	The Evaluation of Community-Based Interventions: Group-Randomization, Limits and Alternatives	6	10/00
207	C. M. Schmidt	Arbeitsmarktpolitische Maßnahmen und ihre Evaluierung: eine Bestandsaufnahme	6	10/00