

# Ist die aktive Arbeitsmarktpolitik in Europa effektiv?

## Ergebnisse einer Meta-Analyse<sup>1</sup>

Viele europäischen Staaten setzen Maßnahmen der sog. Aktiven Arbeitsmarktpolitik – wie z.B. Fortbildungsprogramme oder Lohnsubventionen – zur Bekämpfung der Arbeitslosigkeit ein, oftmals mit erheblichem finanziellen Aufwand. Es gibt jedoch kaum länderübergreifende empirische Erkenntnisse darüber, welche Art von Maßnahme unter welchen Bedingungen für welche Zielgruppe an Teilnehmern tatsächlich wirksam ist. Eine Meta-Analyse auf Basis von Evaluationsstudien aus einzelnen europäischen Ländern kann diese Frage beantworten – mit überraschend eindeutigen Ergebnissen.

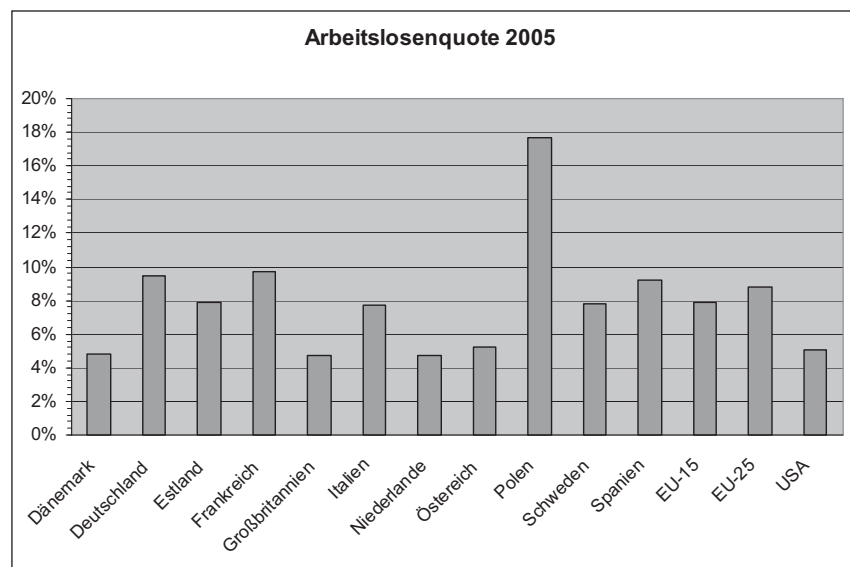
■ Jochen Kluve

Wie Abbildung 1 auf Basis von Daten des Europäischen Statistischen Amtes (Eurostat) für einige ausgewählte EU-Mitgliedsstaaten (und die USA) zeigt, lag im Jahr 2005 die Arbeitslosenquote im Durchschnitt der Staaten der EU-15 bei knapp 8% und in der erweiterten Europäischen Union der EU-25 bei knapp 9%. Während einige Länder von diesem Problem weniger stark betroffen sind, wie beispielsweise Großbritannien, die Niederlande oder Österreich, ist in anderen Ländern die Arbeitslosigkeit jedoch hoch, wie z.B. in Polen oder Frankreich. Auch in Deutschland erreichen die Arbeitslosenzahlen seit Jahren ein Niveau, das die Schaffung von Beschäftigung und die Senkung der Arbeitslosigkeit zu Hauptaufgaben der Politik gemacht hat.

Abbildung 2 verdeutlicht, dass die Bekämpfung der Arbeitslosigkeit in vielen europäischen Staaten keine leichte Aufgabe ist: Die abgebildete Arbeitslosenquote junger Menschen – d.h. von Personen unter 25 Jahren – ist im Jahresmittel sowohl im Durchschnitt der EU-15 (16,8%) als auch im Durchschnitt der EU-25 (18,6%) mehr als doppelt so hoch wie die allgemeine Arbeitslosenquote und liegt in einigen Ländern

*Dr. Jochen Kluve ist Leiter des Kompetenzbereichs „Arbeitsmärkte, Bevölkerung und Gesundheit“ am Rheinisch-Westfälischen Institut für Wirtschaftsforschung (RWI Essen) und insbesondere verantwortlich für „Arbeitsmarktpolitik und Evaluation“.*

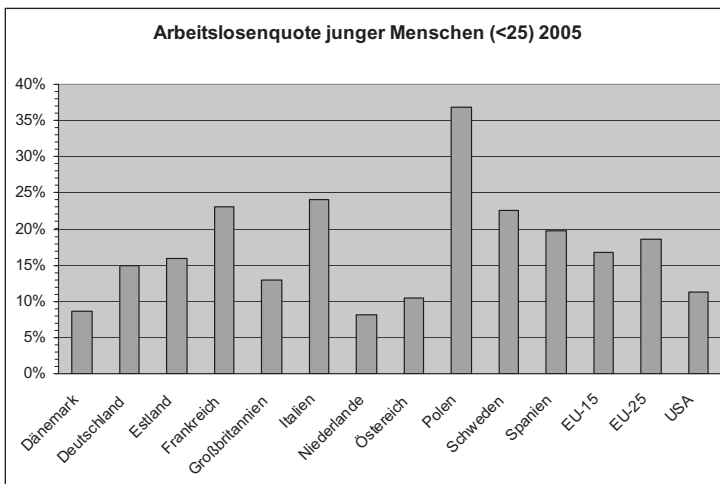
Abbildung 1: Arbeitslosenquote 2005



Quelle: Eurostat

– in der Abbildung z.B. in Frankreich, Italien, Polen – gar deutlich über 20%. Bezieht man in diese Betrachtung noch mit ein, dass nach Angaben der OECD (vgl. „OECD Employment Outlook 2006“) der Anteil der Langzeitarbeitslosen – also jener Menschen, die 12 Monate oder länger ohne Arbeit sind – an allen Arbeitslosen im Jahresmittel 2005 der EU-15 bei 44,3% und dabei in vielen Ländern, wie z.B. Belgien, der Tschechischen Republik, Italien, oder Deutschland, gar über 50% lag, dann wird das Ausmaß der Problematik deutlich. In absoluten Zahlen bedeutet dies, dass im Jahr 2005 in den Staaten der EU-15 14,5 Millionen Menschen arbeitslos waren, darun-

Abbildung 2: Arbeitslosenquote junger menschen (<25) 2005



Quelle: Eurostat

ter 5,9 Millionen Langzeitarbeitslose. In der EU-25 waren dies insgesamt 19 Millionen Menschen ohne Arbeit, darunter 8,5 Millionen Langzeitarbeitslose.

### Der hohen Arbeitslosigkeit in Europa stehen erhebliche Ausgaben für Arbeitsmarktpolitik gegenüber

In der Auseinandersetzung mit diesem Problem der Arbeitslosigkeit ist die Gestaltung der Arbeitsmarktpolitik ein zentrales Element des Handelns der europäischen Regierungen. So wurden nach Angaben von Eurostat im Jahr 2003 in den EU-15 insgesamt 126 Milliarden Euro für die so genannte „Passive Arbeitsmarktpolitik“ – d.h. für Lohnersatzleistungen wie z.B. Arbeitslosengeld – aufgewendet. Darüber hinaus engagieren sich die europäischen Staaten aber insbesondere auch in der so genannten „Aktiven Arbeitsmarktpolitik (AAMP)“, die all jene Politikmaßnahmen umfasst, welche auf die (Wieder-)Eingliederung arbeitsloser Menschen in das Erwerbsleben abzielen. Für derartige Maßnahmen der AAMP gaben die Staaten der EU-15 im Jahr 2003 nach Angaben von Eurostat insgesamt 65 Milliarden Euro aus, was 0,75% ihres Bruttoinlandsproduktes (BIP) entsprach. Abbildung 3 untergliedert dies für einige EU-Mitgliedsstaaten und belegt neben den nennenswerten Ausgaben in einigen Ländern auch eine ausgeprägte Heterogenität zwischen den Ländern: Während beispielsweise Dänemark im Jahr 2004 mehr als 1,5% seines Bruttoinlandsproduktes in Aktive Arbeitsmarktpolitik investierte, lagen Großbritannien oder Estland in ihren diesbezüglichen Ausgaben bei weniger als 0,2%.

Allein aus einer einfachen Gegenüberstellung der immens großen Anzahl an arbeitslosen Menschen in Europa mit den beträchtlichen Ausgaben für Maßnahmen der aktiven Arbeitsmarktpolitik, deren Ziel ja die Verringerung der Arbeitslosigkeit ist, leitet sich unmittelbar die Frage ab, ob diese Ausgaben effektiv sind, d.h. ob es diese Maßnahmen tatsächlich schaffen, Arbeitslosigkeit zu reduzieren. Weitere

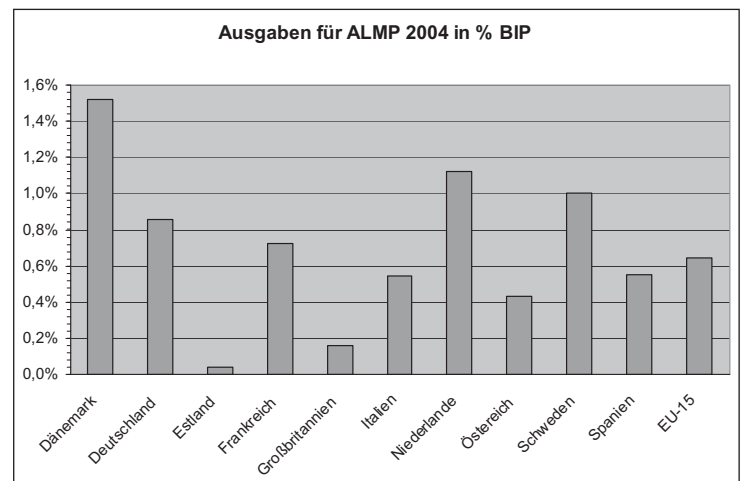
Fragen schließen sich an: Welche der Maßnahmen aktiver Arbeitsmarktpolitik sind es, die Wirkung zeigen? Für welche Teilnehmergruppe(n) entfalten sie eine Wirkung? Und was kann ein Land von den Erfahrungen eines anderen Landes lernen?

### Eine Studie zur Effektivität von AAMP in Europa für die Europäische Kommission

Diese Fragen hat sich auch die Europäische Kommission gestellt und eine Studie zur Wirksamkeit aktiver Arbeitsmarktpolitik in Auftrag gegeben. Diese Studie wurde vom Rheinisch-Westfälischen Institut für Wirtschaftsforschung (RWI Essen) in Kooperation mit einem Netzwerk von Arbeitsmarktforschern aus verschiedenen europäischen Ländern durchgeführt und bildet die Grundlage für den vorliegenden Beitrag.<sup>2</sup> Die Kernaufgabe der Studie war die möglichst stichhaltige Beantwortung der Frage, welche Maßnahme unter welchen institutionellen und ökonomischen Rahmenbedingungen für welche Teilnehmergruppe wirksam ist.

Das methodische Kernproblem bei der Beantwortung dieser Fragestellung ist offensichtlich und besteht in der Tatsache, dass sich einzelne Länder in nahezu jeder relevanten Dimension der Forschungsaufgabe unterscheiden: Verschiedene Länder haben unterschiedliche arbeitsmarktpolitische Strategien, sie wenden aktive Arbeitsmarktpolitik in unterschiedlicher Intensität an, sie zeichnen sich durch unterschiedliche Ausgestaltung spezifischer Maßnahmen aus, und sie befinden sich in unterschiedlichen konjunkturellen Situationen und haben eine unterschiedliche institutionelle Ausgestaltung weiterer arbeitsmarktpolitischer Regelungen. Insbesondere die Rolle der beiden letztgenannten Aspekte – also der ökonomischen und institutionellen Rahmenbedingungen – hinsichtlich der Wirksamkeit von AAMP ist in der Literatur umstritten: Wirkt aktive Arbeitsmarktpolitik eher in Zeiten des konjunkturellen Aufschwungs, wenn generell

Abbildung 3: Ausgaben für ALMP 2004 in % BIP



Quelle: Eurostat

Beschäftigung leichter zu finden ist und dies auch Maßnahmeteilnehmer in gleichem Maße betrifft, oder eher in einer Rezession, wenn die Gruppe der Arbeitslosen insgesamt größer ist und daher eine größere Anzahl an qualifizierten Personen und Maßnahmeteilnehmern enthält, die wiederum leichter Arbeit finden? Und wirkt aktive Arbeitsmarktpolitik eher dann, wenn die Ausgestaltung des Arbeitsmarkts flexibel ist – wie z.B. geringer Kündigungsschutz, flexible Regelungen von Zeitarbeit und befristeter Beschäftigung – oder ist das Gegenteil der Fall?

Der Ansatz der Studie war hierbei, über eine rein deskriptive Betrachtung der vorhandenen Evidenz aus unterschiedlichen Ländern hinauszugehen und zu versuchen, quantitativ robuste und stichhaltige Aussagen zu machen. Dazu wurde eine Meta-Analyse gewählt, eine Methodik, die ihren Ursprung in der medizinischen Forschung hat: hier werden mehrere Studien zur gleichen Forschungsfrage – idealerweise unter gleichen Laborbedingungen – herangezogen, um in der Zusammenschau robustere Evidenz zu erzeugen. Diese Methodik findet auch in der sozialwissenschaftlichen Forschung zunehmend Anwendung. Zwar kann in den Sozialwissenschaften kaum auf Studien zurückgegriffen werden, die alle unter gleichen Bedingungen durchgeführt wurden. Dafür ist es hier aber auch möglich bzw. von Interesse, den Einfluss kontextueller Faktoren – d.h. gerade der unterschiedlichen Rahmenbedingungen – mit zu berücksichtigen.

Im vorliegenden Fall haben wir Studien aus ganz Europa zusammengetragen, die jeweils einzelne Maßnahmen der aktiven Arbeitsmarktpolitik evaluieren: Diese Studien bilden den Datensatz, der Basis der Meta-Analyse ist. Die quantitative Analyse besteht dann darin, die Tatsache, ob eine Evaluationsstudie einen positiven Effekt der untersuchten Maßnahme findet, mit einer ganzen Reihe erklärender Variablen zu korrelieren: diese erklärenden Variablen umfassen insbesondere die Art der Maßnahme, die Zielgruppe, die wirtschaftliche Situation des Landes zum Zeitpunkt der Maßnahme, sowie die arbeitsmarktrechtliche Ausgestaltung des Landes zum Zeitpunkt der Maßnahme. So lässt sich feststellen, ob und mit welcher Intensität diese Einflussfaktoren mit der Wirksamkeit von AAMP in Zusammenhang stehen.

### **Erster Analyserahmen: Die Europäische Beschäftigungsstrategie**

Die quantitative Analyse ist eingebettet in drei Rahmen, a) der Europäischen Beschäftigungsstrategie, b) der adäquaten Klassifizierung von Maßnahmen, und c) der Methodik der Evaluation einzelner Maßnahmen. Diese drei Analyserahmen ergeben sich für unsere Studie aus den drei Tatsachen, dass erstens die Europäische Beschäftigungsstrategie Leitlinien für die Arbeitsmarktpolitik der EU-Mitgliedstaaten entwirft und somit die aktive Arbeitsmarktpolitik und deren Evaluation beeinflusst, dass zweitens eine länderübergreifende Analyse von AAMP hinreichend homogene Maßnahmentypen vergleichen muss, und dass drittens die einzelnen Evaluationsstudien, auf denen wir aufbauen, methodisch überzeugende Evidenz liefern müssen.

Aktive Arbeitsmarktpolitik ist ein wichtiges Element der *Europäischen Beschäftigungsstrategie*. Diese wurde im Jahr 1997 im Rahmen des *Amsterdamer Vertrags* als so genannter „Luxemburger Prozess“ begonnen, als die Europäische Kommission angesichts unbefriedigender Entwicklung der Arbeitsmärkte ihrer Mitgliedsstaaten der Beschäftigungspolitik erstmals den gleichen Status zuschrieb wie anderen zentralen Zielen der Wirtschaftspolitik der Europäischen Union. Artikel 2 des Amsterdamer Vertrags beispielsweise formuliert, dass die Mitgliedsstaaten die Förderung von Beschäftigung als gemeinsame Angelegenheit betrachten und ihr Handeln miteinander abstimmen sollen. Artikel 3 erkennt formell an, dass eine hohe Beschäftigungsquote ein explizites Ziel in der Formulierung und Umsetzung von Politikmaßnahmen der Europäischen Gemeinschaft sein sollen.

Dementsprechend wurden in der Entwicklung der Europäischen Beschäftigungsstrategie konkrete *Beschäftigungsziele* formuliert. Das Treffen des Europäischen Rates in Lissabon im Jahr 2000 legte fest, dass bis zum Jahr 2010 die Europäische Union die Voraussetzungen für Vollbeschäftigung wiedererlangt und die Soziale Kohäsion verstärkt haben soll. Insbesondere sollen bis 2010 die durchschnittliche Gesamtbeschäftigungsrate in der EU auf 70% sowie jene der Frauen auf über 60% erhöht werden. Beim Treffen in Stockholm 2001 legte der Europäische Rat Zwischenziele fest (67% Gesamtbeschäftigungsrate und 57% für Frauen bis 2005). Das Treffen in Barcelona 2002 bekräftigte, dass Vollbeschäftigung das übergreifende Ziel ist, und forderte eine verstärkte Beschäftigungsstrategie ein, um die Ziele von Lissabon in einer erweiterten Europäischen Union zu untermauern. Im Jahr 2005 lag nach Angaben von Eurostat die Gesamtbeschäftigungsquote in den Staaten der EU-15 bei 65,2% (Frauen: 57,4%) und in der gesamten EU-25 bei 63,8% (Frauen: 56,3%). Auch wenn die entsprechenden Zwischenziele damit nicht ganz erreicht wurden, unterliegt den Beschäftigungsquoten ein stetig ansteigender Trend über die letzten Jahre.

Die Europäische Beschäftigungsstrategie ist in ihrer Umsetzung von einem Feedback-Prozess gekennzeichnet: Die Europäische Kommission formuliert jährlich *beschäftigungspolitische Leitlinien*, auf deren Basis die Mitgliedsstaaten so genannte *Nationale Aktionspläne* (NAPs) entwerfen. Unter Berücksichtigung der tatsächlichen Entwicklung und der NAPs entsteht dann ein jährlicher *Gemeinsamer Beschäftigungsbericht* (Joint Employment Report), der wiederum Grundlage für die nächsten beschäftigungspolitischen Leitlinien ist, usw. Diese ursprüngliche Struktur wurde im Jahr 2005 dahingehend geändert, dass die beschäftigungspolitischen Leitlinien nunmehr als *integrierte Leitlinien* gemeinsam mit den makroökonomischen und mikroökonomischen Leitlinien entworfen und nur alle drei Jahre von der Kommission vollständig überprüft werden.

Neben den Beschäftigungszielen ist es ein wichtiger Effekt der Europäischen Beschäftigungsstrategie, dass sie die Evaluationsbemühungen in den EU-Mitgliedstaaten maßgeblich vorangebracht hat.

## Zweiter Analyserahmen: Die Kategorisierung von Maßnahmen

Um die aktive Arbeitsmarktpolitik in einer länderübergreifenden Perspektive vergleichbar zu machen, war es auch eine Kernaufgabe der Studie, die unterschiedlichen Maßnahmentypen einzelner Länder in hinreichend homogene Kategorien zusammenzufassen. Hierbei haben wir uns an entsprechenden Klassifikationen orientiert, wie sie bereits von der OECD und Eurostat verwendet werden. Allein die vierte Kategorie (vgl. unten) weicht von den bekannten Klassifikationen ab.

Die erste Kategorie der Maßnahmen aktiver Arbeitsmarktpolitik bilden *Fortbildungs- und Qualifizierungsmaßnahmen*. Dies sind im Kern alle Maßnahmen, die das Humankapital und die Produktivität der Teilnehmer – und damit ihre Beschäftigungswahrscheinlichkeit – erhöhen sollen. Diese Kategorie umfasst allgemeine Fortbildung ebenso wie berufliche Ausbildung. Derartige Bildungsmaßnahmen sind gewissermaßen die „klassische“ Maßnahme der aktiven Arbeitsmarktpolitik.

Die zweite Kategorie von AAMP sind so genannte *Anreizprogramme im privaten Sektor*. Dies umfasst all jene Maßnahmen, die (finanzielle) Anreize schaffen, um das Marktverhalten von Arbeitnehmern und / oder von Arbeitgebern zu verändern. Die bekannteste Maßnahme in dieser Kategorie sind Lohnsubventionen. Auch Existenzgründungsprogramme gehören zu dieser Kategorie.

In der dritten Gruppe finden sich Maßnahmen der direkten Beschäftigungsschaffung im öffentlichen Sektor, also der Bereitstellung öffentlicher Arbeitsplätze oder Aktivitäten, die öffentliche Güter und Dienstleistungen produzieren. Diese Maßnahmen sind in der Regel an jene Arbeitslosen mit den größten Vermittlungsschwierigkeiten gerichtet, im Versuch, diese mit dem Arbeitsmarkt in Kontakt zu halten und den Verlust an Humankapital über längere Phasen der Arbeitslosigkeit zu vermeiden. Die so geschaffene Beschäftigung zeichnet sich jedoch häufig nicht durch ihre Nähe zum ersten Arbeitsmarkt aus.

Die vierte Kategorie wurde von uns als *Dienstleistungen und Sanktionen* bezeichnet und umfasst alle Maßnahmen, die auf eine Steigerung der Vermittlungseffizienz ausgerichtet sind. Damit erweitern wir die üblicherweise verwendete Kategorie „Vermittlungshilfe“, insbesondere durch Einbezug von Sanktionen. Letzteres reflektiert die Tatsache, dass in den EU-Mitgliedstaaten die Dienstleistung „Arbeitsvermittlung“ zunehmend an Anforderungen an die Suchaktivität der arbeitslosen Person geknüpft ist, was sich z.B. in Deutschland im Konzept des „Fördern und Fordern“ spiegelt. Die gemeinsame Kategorisierung erscheint gerechtfertigt, da all diese Maßnahmen – wie Bewerbungshilfe, Monitoring, Beratung, Sperrzeiten – ein gemeinsames Ziel verfolgen: alle zielen auf die Erhöhung der Effizienz der Beschäftigungssuche ab. Während prinzipiell in den meisten europäischen Ländern sowohl private als auch öffentliche Vermittler tätig sind, werden diese Aufgaben zum Großteil von der öffentlichen Arbeitsvermittlung geleistet.

Zusätzlich zu den genannten Kategorien zeichnen sich viele Maßnahmen der AAMP durch den Zuschnitt auf bestimmte Zielgruppen aus. Erwähnenswert sind hier insbesondere Programme speziell für junge Arbeitslose – derartige Maßnahmen gibt es in fast allen europäischen Ländern, entsprechend der oben dargestellten verbreiteten Problematik der Jugendarbeitslosigkeit. Auch aktive Maßnahmen speziell für Menschen mit Behinderung gibt es in vielen europäischen Ländern.

## Dritter Analyserahmen: Die Evaluationsmethodik

Unser oben beschriebener Ansatz einer länderübergreifenden Meta-Analyse muss notwendigerweise auf stichhaltige und verlässliche Evidenz aus einzelnen Ländern zurückgreifen. Hierzu wurde, wie erwähnt, auf Studien zurückgegriffen, die einzelne AAMP-Programme evaluieren. Alle verwendeten Studien sind akademische ökonomische Analysen, die versuchen, methodisch überzeugend das so genannte *Evaluationsproblem* zu lösen. Dieses Problem, mit dem sich jede Analyse, deren Ziel die Ermittlung eines kausalen Effektes einer Maßnahme ist, auseinandersetzen muss, wird in beistehendem Kasten in Grundzügen erläutert.

Vor dem Hintergrund dieser drei Analyserahmen – der beschäftigungspolitischen Rolle von AAMP in Europa, der länderübergreifenden Einordnung von AAMP-Maßnahmen, sowie der überzeugenden Evaluation von AAMP – lässt sich die Situation in europäischen Ländern genauer betrachten. Hierzu fokussiert unsere Studie zuerst auf einer Reihe von Länderberichten aus ausgewählten EU-Mitgliedsstaaten, bevor im zweiten Fokus die Meta-Analyse versucht, quantitative Aussagen zur Wirksamkeit von AAMP zu machen.

## Die Situation in einzelnen Ländern

In die genauere Länderbetrachtung wurden Österreich, Dänemark, Estland, Frankreich, Deutschland, Italien, die Niederlande, Polen, Spanien, Schweden und Großbritannien einbezogen. Die Länderstudien folgen hierbei aus Gründen der Vergleichbarkeit einer einheitlichen Struktur und diskutieren jeweils (a) die wirtschaftliche Situation, (b) die Institutionen des Arbeitsmarktes, (c) den Einsatz von AAMP, und (d) die Evaluationen von AAMP. Es ist nicht überraschend, dass sowohl die wirtschaftliche Situation als auch die institutionelle Ausgestaltung von Land zu Land sehr unterschiedlich ist, und von momentan leistungsfähigen (z.B. Dänemark, Estland) zu eher stagnierenden Ökonomien (z.B. Deutschland) schwankt, sowie von eher flexiblen (z.B. Großbritannien) zu eher stark regulierten Arbeitsmärkten (z.B. Frankreich, Deutschland). Einige Länder verwenden einen relativ großen Anteil ihres Bruttoinlandsproduktes für AAMP (z.B. die Niederlande, Dänemark, Schweden, vgl. auch Abbildung 3 oben) und haben ein äußerst umfangreiches Portfolio an Maßnahmen (z.B. Deutschland), während andere Länder deutlich weniger ausgeben (z.B. Großbritannien, Italien) und eine relative geringe Zahl an Programmen einsetzen (z.B. Estland, Spanien). Dänemark hat zweifelsohne die umfassendste AAMP-Strategie mit

erheblichem Bemühen, alle arbeitslosen Personen zu aktivieren.

Ähnlich der Unterschiede im Einsatz von AAMP variieren auch die Evaluationsbemühungen zwischen einzelnen Ländern. Schweden ist bekannt für seine lange Tradition von AAMP und deren Evaluation, und die Niederlande und Großbritannien heben sich durch die Verwendung einzelner randomisierter Experimente ab. In Spanien und Italien beispielsweise ist kaum von einer „Evaluationskultur“ zu sprechen, was vermutlich mit einem noch eingeschränkten und erst beginnenden Einsatz von AAMP zusammenhängt. Deutschland ist ein Beispiel für ein Land, in dem – trotz einer relativ langen Tradition mit AAMP – die Evaluation von arbeitsmarktpolitischen Maßnahmen bis vor wenigen Jahren fast nicht existierte, und in dem sich dann eine Evaluationskultur sehr schnell entwickelte.

### Meta-Analyse: Idee und Technik

Angesichts dieser Heterogenität zwischen europäischen Ländern war es, wie eingangs erwähnt, Kernaufgabe unserer Studie, einen methodischen Zugang zu einer länderübergreifenden Betrachtung der Wirksamkeit von AAMP zu finden. Hierzu wählten wir die Technik der Meta-Analyse, d.h. eine systematische Analyse einzelner Studien zur Effektivität aktiver Arbeitsmarktpolitik. Grundlage einer solchen Meta-Analyse ist die Datensammlung dieser einzelnen Studien, die einem so genannten „Protokoll“ folgt (für ausführlichere Informationen zu Meta-Analysen im sozialwissenschaftlichen Kontext vgl. [www.campbellcollaboration.org](http://www.campbellcollaboration.org)).

Gemäß den Richtlinien unseres Protokolls wurden nur Studien einbezogen, die folgende Bedingungen erfüllen:

- a) Empirische Studien, die für die Selektion in Teilnahme- und Kontrollgruppe kontrollieren;
- b) Studien, die einzelne Maßnahmen betrachten und Maßnahmen nicht bündeln;
- c) Studien, die Maßnahmeeffekte relativ zur Nichtteilnahme angeben und nicht relativ zu anderen Maßnahmen;
- d) Mikroökonomische Studien, d.h. Studien, die Effekte auf individueller Ebene ausweisen.

Neben diesen Restriktionen bestand natürlich ein Interesse daran, möglichst viele Studien einzubeziehen, um einen hinreichend großen Datensatz für eine quantitative Untersuchung zu erhalten.

### Ein Datensatz aus 137 Einzelstudien

Insgesamt ist es uns gelungen, 137 Evaluationen einzelner AAMP-Maßnahmen aus europäischen Ländern (EU-Mitgliedstaaten sowie Norwegen und die Schweiz) zusammenzutragen. Von diesen 137 Evaluationen ermitteln 75 einen positiven Effekt auf die individuelle Beschäftigungswahrscheinlichkeit, 33 finden keinen signifikanten Effekt und 29 berichten von einer negativen

Wirksamkeit der Maßnahme. Diese Information zur Wirksamkeit jeder einzelnen Maßnahme ist die zentrale Ergebnisgröße, die in der quantitativen Analyse mit einer Reihe an erklärenden Variablen korreliert wird.

### Welche Faktoren könnten die Wirksamkeit von AAMP beeinflussen?

Die Liste der erklärenden Variablen – also jener Variablen, von denen wir vermuten, dass sie die ermittelte Wirksamkeit einer Maßnahme beeinflussen könnten – kann in vier Gruppen unterteilt werden: i) die Art der Maßnahme, ii) das Studiendesign, iii) den institutionellen Kontext, iv) den ökonomischen Kontext. Die erste Gruppe – Art der Maßnahme – umfasst genau jene Kategorien, die bereits oben in der Klassifizierung der Maßnahmetypen beschrieben wurden. In den Daten finden sich hierzu genau 70 Fortbildungsmaßnahmen, 23 Anreizprogramme im privaten Sektor, 26 öffentliche Beschäftigungsprogramme und 21 Maßnahmen der „Dienstleistungen und Sanktionen“ (diese Zahlen summieren sich zu 140 anstelle von 137, da 3 der Anreizprogramme öffentlichen und privaten Sektor mischen und keine diesbezügliche Differenzierung möglich war). 35 Evaluationen betreffen Maßnahmen, die speziell auf jugendliche Arbeitslose zugeschnitten sind.

In der zweiten Gruppe, Studiendesign, wird erfasst, ob die Maßnahme experimentell oder nicht-experimentell evaluiert wurde, wann sie evaluiert wurde (d.h. in welcher Dekade), und wie groß die Stichprobe ist, auf der die Schätzung der Wirksamkeit beruht. Hierzu würde man annehmen, dass Experimente besonders stichhaltige Evidenz generieren, dass jüngere Evaluationsstudien auf ausführlicheren Daten beruhen und allgemein anerkannte Methoden verwenden (deren Entwicklung in der Evaluationsforschung seit Mitte der 1990er Jahre einen immensen Fortschritt gemacht hat), und dass größere Stichproben Effekte präziser messen können. In den Daten zeigt sich, dass nur 9 der Evaluationen experimentell durchgeführt wurden. 4 der Studien stammen noch aus den 1970er Jahren, 36 aus den 1980ern, die große Mehrzahl mit 81 aus den 1990ern, und 16 aus dem Zeitraum seit 2000. 43% der Studien basieren auf kleinen Stichproben (d.h. die Anzahl N der Beobachtungen in Teilnahme- und Kontrollgruppe zusammen ist geringer als 1000), 40% basieren auf mittelgroßen Stichproben ( $1000 \leq N < 10000$ ), und 17% auf großen Stichproben ( $N > 10000$ ).

Die dritte Gruppe der erklärenden Variablen umfasst vier Indikatoren, die versuchen, den institutionellen Kontext abzubilden, in dem eine Maßnahme umgesetzt wurde und der potenziell die Effektivität beeinflusst. Diese Indikatoren werden immer in einem bestimmten Land zu genau jenem Zeitpunkt gemessen, zu dem auch die betrachtete AAMP-Maßnahme durchgeführt wurde, d.h. sie erfasst für jedes Land auch Variation über die Zeit, z.B. wenn ein Land neue Regelungen zur Flexibilisierung des Arbeitsmarktes durchgeführt hat. Der institutionelle Kontext betrifft hier insbesondere Regulierungen, die potenziell die Bereitschaft von Arbeitgebern beeinflussen,

Teilnehmer aus AAMP-Maßnahmen einzustellen, sowie die Bereitschaft von Arbeitnehmern, Stellenangebote anzunehmen. Letztere wird durch die Lohnersatzrate abgebildet. In den Daten liegt diese zwischen 17,5% (Großbritannien in den späten 1990ern) und 63,7% (Dänemark 1996). Die anderen drei Variablen sind Indizes für die Striktheit der Regelungen hinsichtlich Kündigungsschutz, befristeter Beschäftigung und Zeitarbeit, entnommen Angaben der OECD („OECD Employment Outlook 2004“). Der Kündigungsschutzindex nimmt hierbei Werte von 0,8 (Großbritannien in den frühen 1980ern) bis 4,3 (Portugal in den späten 1990ern) an. Der Index der befristeten Beschäftigung reicht von 0 (mehrere Länder inklusive Großbritannien) bis 5,5 (Belgien in den frühen 1990ern), während der Regulierungsindex der Zeitarbeit von 0,5 (mehrere Länder inklusive Dänemark) bis 5,5 (Schweden im Zeitraum 1970er bis frühe 1990er) reicht.

Die vierte Variablengruppe schließlich gibt die wirtschaftliche Situation an, in Verhältnis zu der wir die in den Studien identifizierten Effekte interpretieren möchten. Auch diese Variablen werden wiederum zu jenem Zeitpunkt (bzw. im Durchschnitt über jenen Zeitraum hinweg) gemessen, zu dem die evaluierte AAMP-Maßnahme durchgeführt wurde. Die wirtschaftlichen Kontextvariablen umfassen die Arbeitslosenquote, die jährliche Wachstumsrate des BIP, sowie die Ausgaben für AAMP in Prozent des BIP. In unseren Daten liegen die Arbeitslosenquoten zwischen 1,9% (Schweden in den späten 1970ern) und 16,5% (Irland in den späten 1980ern). Die jährliche Wachstumsrate des BIP reicht von -0,7 (Finnland im Zeitraum 1990-1995) bis +7,1 (Estland 2000-2002). Der Ausgabenindex für AAMP liegt zwischen 0,03% des BIP (Slowakische Republik 1993-1998) und 2,68% des BIP (Schweden in den frühen 1990ern).

### Überraschend eindeutige Ergebnisse

In der quantitativen Analyse korreliert man die Ergebnisgröße – also die Tatsache, ob eine Evaluationsstudie einen positiven, einen negativen, oder gar keinen Effekt der Maßnahme auf die individuelle Beschäftigungswahrscheinlichkeit attestiert – mit der im vorigen Abschnitt beschriebenen Liste an erklärenden Variablen. Die empirischen Ergebnisse sind ebenso überraschend deutlich wie auch über verschiedene Spezifikationen der Regression hinweg robust.

Zum einen zeigen die Ergebnisse, dass es kaum einen Zusammenhang zwischen institutionellen oder ökonomischen Kontextfaktoren und der Wirksamkeit aktiver Arbeitsmarktpolitik gibt. Einzige Ausnahme ist der Kündigungsschutz: Dieser scheint einen wichtigen systematischen Effekt auf die Wirksamkeit einer Maßnahme zu haben, dergestalt, dass AAMP in Ländern mit flexiblerem, d.h. geringerem Kündigungsschutz mit höherer Wahrscheinlichkeit positive Beschäftigungswirkungen hat. Dies mag damit zu erklären sein, dass ein weniger restriktiver Kündigungsschutz die Eintrittsschwelle in den ersten Arbeitsmarkt senkt, und sich dies insbesondere für Teilnehmer an Maßnahmen der aktiven Ar-

beitsmarktpolitik positiv auswirkt. Überraschend ist an den Ergebnissen insbesondere, dass keiner der anderen kontextuellen Faktoren – also weder Lohnersatzrate noch Arbeitslosenquote noch Wachstumsrate des BIP – einen signifikanten Zusammenhang mit der Wirkung von AAMP erkennen lässt.

Zum anderen zeigen die Ergebnisse, dass für die Wirksamkeit einer Maßnahme die Art der Maßnahme von ganz entscheidender Bedeutung zu sein scheint: Traditionelle Fortbildungs- und Qualifizierungsmaßnahmen zeigen eine geringfügige Wahrscheinlichkeit, einen positiven Effekt auf die Beschäftigungsrate nach der Teilnahme zu haben (etwa die Hälfte aller Qualifizierungsmaßnahmen zeigt positive Wirkungen). Relativ dazu weisen Anreizprogramme im privaten Sektor sowie Dienstleistungen und Sanktionen signifikant bessere Effekte auf. Deren Wahrscheinlichkeit, einen positiven Effekt zu messen, ist etwa 40 Prozentpunkte höher als bei Fortbildungs- und Qualifizierungsmaßnahmen. Im Vergleich dazu ist für Beschäftigungsprogramme im öffentlichen Sektor die Wahrscheinlichkeit eines positiven Maßnahmeeffekts signifikant 30-40 Prozentpunkte geringer. Und auch die Zielgruppe der Maßnahme scheint eine Rolle zu spielen: Maßnahmen für junge Arbeitslose schneiden deutlich schlechter ab als Maßnahmen für Erwachsene, und weisen eine um 40-60 Prozentpunkte geringere Wahrscheinlichkeit auf, einen positiven Maßnahmeeffekt zu messen.

### Schlussfolgerungen

Die Schlussfolgerungen aus diesen empirischen Befunden sind recht nahe liegend. Die politischen Entscheidungsträger sollten bei der Entwicklung des AAMP-Portfolios insbesondere den Programmtyp berücksichtigen: Fortbildungs- und Qualifizierungsmaßnahmen sollten fortgesetzt werden, und Anreizprogramme im privaten Sektor sollten gefördert und ausgebaut werden. Besondere Aufmerksamkeit sollte auf Dienstleistungen und Sanktionen gerichtet werden, ein viel versprechender und durch relativ geringe Ausgaben kosten-effektiver Maßnahmetyp. Ein gut balanciertes System aus grundlegenden Dienstleistungen wie Unterstützung bei der Stellensuche und Beratung und Monitoring, zusammen mit geeigneten Sanktionen bei Nichtbefolgung, scheint einen wesentlichen Beitrag zur Verbesserung der Effizienz der Beschäftigungssuche beitragen zu können. Dies ist beispielsweise der Ansatz des „New Deal“ in Großbritannien.

Beschäftigungsprogramme im öffentlichen Sektor dagegen sind selten wirksam und gelegentlich gar schädlich hinsichtlich der Beschäftigungswahrscheinlichkeit der Teilnehmer. Gemäß dieser Resultate sollten sie nicht fortgesetzt werden, außer es gäbe andere Begründungen für diese Art der Förderung: Einige Länder haben das Ziel solcher Beschäftigungsprogramme bereits umformuliert, dergestalt, dass diese die „Beschäftigungsfähigkeit“ anstelle tatsächlicher Beschäftigung fördern oder erhalten sollen. Dies ist jedoch eine Ergebnisvariable, die empirisch sehr schwer messbar ist.

Jungen arbeitslosen Menschen scheint mit aktiven Maßnahmen nur schwer zu helfen zu sein. Es bleibt unklar, ob aus diesem enttäuschenden Resultat der Schluss folgt, dass diese Programme abgeschafft werden sollten, oder aber, dass diese Programme neu gestaltet werden und besonders viel Aufmerksamkeit erhalten sollten. Es könnte auch der Fall sein, dass aktive Arbeitsmarktpolitik gar nicht die geeignete Maßnahme für diese Zielgruppe darstellt, und die Politik sich eher auf präventive Maßnahmen konzentrieren sollte, die Jugendliche und junge Erwachsene erst gar nicht zu Benachteiligten auf dem Arbeitsmarkt machen.

Die Entwicklung einer „Evaluationskultur“ zeigt sich in allen europäischen Ländern positiv, auch wenn sich natürlich unterschiedliche Länder auf unterschiedlichen Stufen dieser Entwicklung befinden. Eine offensichtliche Schlussfolgerung unserer Studie ist, dass Evaluationsbemühungen fortgesetzt und ausgebaut werden sollten. Eine weiter verfeinerte Meta-Analyse auf Basis einer

weiter ausgebauten Datenbasis europäischer Evaluationsstudien würde weiterhin wichtige Erkenntnisse zur Wirksamkeit von AAMP liefern können, insbesondere da sich Datenqualität und Methodologie erwartungsgemäß weiter verbessern werden. Ungeachtet der beachtlichen Fortschritte der nicht-experimentellen Evaluationsmethodik sollten dennoch mehr Mitgliedsstaaten mit Interesse an der Wirksamkeit ihrer Maßnahmen den Einsatz randomisierter Experimente in Erwägung ziehen, angesichts der großen Aussagekraft ihrer Ergebnisse.

### Fußnoten

- 1 Dieser Beitrag ist eine überarbeitete Fassung des Artikels „Die Wirksamkeit aktiver Arbeitsmarktpolitik in Europa“, erschienen im Bundesarbeitsblatt 10-2006.
- 2 Vgl. hierzu Kluge, J., D. Card, M. Fertig, M. Gora, L. Jacobi, P. Jensen, R. Leetmaa, L. Nima, E. Patacchini, S. Schaffner, C.M. Schmidt, B. van der Klaauw und A. Weber (2007), *Active Labor Market Policy in Europe: Performance and Perspectives*, Springer-Verlag: Berlin et al., sowie Kluge, J. (2007), „The effectiveness of European Active Labor Market Policy“, überarbeitete Fassung von IZA Disc.Paper 2018.

## Das Evaluationsproblem

Um die Wirksamkeit einer Maßnahme zu bestimmen, betrachtet das Kausalmodell der Evaluationsforschung eine *Beobachtungseinheit* (wie zum Beispiel eine Person), die an einer *Maßnahme* (wie zum Beispiel einem Fortbildungsprogramm) teilnehmen kann oder auch nicht. Dies wird im Modell für jede Person durch einen sog. *Teilnahmeindikator* angezeigt:  $D=1$  entspricht der Teilnahme,  $D=0$  entspricht der Nichtteilnahme. Von zentralem Interesse ist der Effekt der Maßnahme auf eine nach der Maßnahme gemessene *Ergebnisvariable*  $Y$  (wie z.B. die Beschäftigungswahrscheinlichkeit der Person). Diese kann für jede Person die beiden potentiellen *Ergebnisse*  $Y_1$  (im Falle der Teilnahme, also  $D=1$ ) oder  $Y_0$  (im Falle der Nichtteilnahme,  $D=0$ ) annehmen. Die Differenz  $\Delta$  aus Teilnahmeergebnis und Nichtteilnahmeergebnis misst den kausalen Effekt der Maßnahme:  $\Delta=Y_1-Y_0$ .

Da für die einzelne Beobachtungseinheit immer nur einer der beiden Fälle tatsächlich eintreten kann (entweder Teilnahme oder Nichtteilnahme) und somit auch nur eines der beiden potentiellen Ergebnisse eintreten kann, ist das jeweils andere potentielle Ergebnis (d.h. das Teilnahmeergebnis eines Nichtteilnehmers und das Nichtteilnahmeergebnis eines Teilnehmers) eine nicht in Daten beobachtbare, sog. *kontrafaktische Größe*. Der zentrale konzeptionelle Punkt des Modells ist somit, dass jedes Individuum zwei potentielle Ergebnisse aufweist, deren Vergleich den kausalen Effekt messen würde, die aber nicht gleichzeitig in Daten beobachtbar sind, da ja für jede Person nur ein Ergebnis eintritt. Dies wird als das *Evaluationsproblem* bezeichnet.

Da Effekte auf individueller Ebene somit nicht direkt messbar sind, betrachtet man in der empirischen Evaluationsforschung in der Regel Mittelwerte, die das durchschnittliche Ergebnis der Teilnehmer mit dem durchschnittlichen Ergebnis der Nichtteilnehmer vergleicht. Hierbei ist es von zentraler Bedeutung, eine wirklich vergleichbare Kontrollgruppe zu identifizieren, da diese ja eine Aussage darüber erlauben muss, was mit den Teilnehmern im Falle der Nichtteilnahme hinsichtlich der Ergebnisgröße geschehen wäre. Diese Vorgehensweise der Evaluation erlaubt dann eine Abschätzung des so genannten „Durchschnittlichen Effekts der Teilnahme auf die Teilnehmer“ (Average Treatment Effect on the Treated, ATET).

Da es für die Validität einer Evaluationsstudie somit entscheidend ist, ob die Kontrollgruppe der Teilnahmegruppe in allen relevanten Größen so ähnlich wie möglich ist, wäre dies prinzipiell am überzeugendsten durch ein randomisiertes Experiment – also der zufälligen Zuordnung aller potentiellen Teilnehmer in Teilnahme- und Kontrollgruppe – zu lösen, wie es in der medizinischen Forschung üblich ist und z.B. in den USA auch zur Evaluation arbeitsmarktpolitischer Programme bereits verwendet wurde. Dieses Vorgehen gewährleistet, dass sich Teilnahme- und Kontrollgruppe im Durchschnitt weder in beobachtbaren noch in unbeobachtbaren Charakteristika unterscheiden. Nicht-experimentelle Ansätze, wie sie in der europäischen Arbeitsmarktforschung üblich sind, bilden in der Regel eine Kontrollgruppe *ex post*, und können daher nur für messbare Unterschiede – Alter, Geschlecht, Bildungsstand, etc – kontrollieren, nicht aber für potentiell relevante unbeobachtbare Charakteristika wie beispielsweise Motivation.