

IAB *Forschungsbericht*

Ergebnisse aus der Projektarbeit des Instituts für Arbeitsmarkt- und Berufsforschung

Nr. 3/2005

Vermittlungsgutscheine

Zwischenergebnisse der Begleitforschung 2004

Teil III

Mikroökonomische Wirkungsanalyse

Anja Heinze, Friedhelm Pfeiffer, Alexander Spermann, Henrik Winterhager

Vermittlungsgutscheine

Zwischenergebnisse der Begleitforschung 2004

Teil III

Mikroökonomische Wirkungsanalyse

Anja Heinze, Friedhelm Pfeiffer, Alexander Spermann, Henrik Winterhager (ZEW Mannheim)

Mit der Publikation von Forschungsberichten will das IAB der Fachöffentlichkeit Einblick in seine laufenden Arbeiten geben. Die Berichte sollen aber auch den Forscherinnen und Forschern einen unkomplizierten und raschen Zugang zum Markt verschaffen. Vor allem längere Zwischen- aber auch Endberichte aus der empirischen Projektarbeit bilden die Basis der Reihe, die den bisherigen „IAB-Werkstattbericht“ ablöst.

Inhaltsverzeichnis

Abstract.....	4
Vorbemerkung.....	5
1 Einführung	6
2 Methodische Überlegungen.....	7
2.1 Individualeffekte versus gesamtwirtschaftliche Effekte	8
2.2 Das mikroökonomische Evaluationsproblem.....	9
2.3 Drei grundsätzliche Ansätze zur Lösung des Evaluationsproblems ..	13
2.4 Vergleich von Teilnehmern und Nichtteilnehmern: Matching	15
2.4.1 Annahmen	15
2.4.2 Drei Varianten des Matching	17
2.4.2.1 Exaktes Matching	17
2.4.2.2 Propensity Score Matching	17
2.4.2.3 Mahalanobis Matching	23
2.4.3 Kritische Reflexion des Matching-Ansatzes	23
2.4.4 Diskussion der Matching-Annahmen für vorliegende Untersuchung	26
2.5 Methodische Details.....	30
2.5.1 Datum des Maßnahmebeginns und zeitabhängige Variablen	30
2.5.2 Balancing Tests	35
3 Datenbasis	36
4 Ergebnisse	38
4.1 Schätzung des Propensity Scores	39
4.2 Qualität des Matchings: Balancing Tests.....	51
4.3 Schätzung der Average Treatment Effekte für die Teilnehmer	56
4.4 Gruppenspezifische Treatment Effekte	59
4.5 Sensitivitätsanalysen	77
5 Zusammenfassung	78
Literatur	79

Abstract

Aufbauend auf den Mikrodaten werden in diesem Forschungsbericht die Determinanten der Ausgabe von Vermittlungsgutscheinen und deren direkte Wirkungen auf der Individualebene untersucht. Die Analysen legen den Schluss nahe, dass die Gruppe der Arbeitslosen, die einen Vermittlungsgutschein erhalten haben, keine Zufallsauswahl aus der Gruppe aller Arbeitslosen ist. Es scheinen somit Selektionsmechanismen zu existieren, die auch für den Erfolg der Vermittlungsgutscheine relevant sind. In der Mikroanalyse wird die Selektivität der Ausgabe berücksichtigt. Dazu werden nicht-parametrische Matching-Verfahren verwendet, die auf der Annahme der bedingten Unabhängigkeit (CIA) beruhen. Für die Vermittlungsgutscheine, die im Mai und Juni 2003 ausgegeben wurden, wird die Wahrscheinlichkeit der Einmündung in eine reguläre Beschäftigung innerhalb von sechs Monaten abgeschätzt. Der für die Gruppe der Arbeitslosen, die einen Vermittlungsgutschein erhalten haben, geschätzte durchschnittliche Maßnahmeeffekt beträgt den Ergebnissen zu Folge 4,8 Prozentpunkte in Ostdeutschland und 3,7 Prozentpunkte in Westdeutschland. Die ermittelten Maßnahmeeffekte sind für Arbeitslose mit kurzer Arbeitslosigkeitsdauer, deren Wert der Vermittlungsgutscheine bei 1.500 € liegt, höher als für Arbeitslose mit längerer Arbeitslosigkeitsdauer, deren Wert der Vermittlungsgutscheine bis zu 2.500 € betragen kann.

Vorbemerkung

Im Rahmen des IAB-Auftragsprojektes „Begleitforschung zur Vermittlung: Implementation und Evaluierung der Beauftragungen nach § 37(a) SGB III, Vermittlungsgutscheinen (VGS) und Personalserviceagenturen (PSA)“ war für das Jahr 2004 von den Auftragnehmern sinus München und ZEW Mannheim als Schwerpunkt der Vermittlungsgutschein zu evaluieren. Mit dem zweiten Zwischenbericht liegen aussagefähige Zwischenergebnisse zur Implementation des Instruments, zur Einschaltung privater Arbeitsvermittler und zur arbeitsmarktlichen Wirksamkeit der Vermittlungsgutscheine vor. Wegen seines Umfangs wird der Zwischenbericht in sieben Teilen veröffentlicht. Teil I beschreibt das Projektmonitoring mit den verwendeten Daten und liefert deskriptive Auswertungen. Teil II enthält eine Typisierung der Arbeitsagenturen im Hinblick auf die Strategien und Vorgehensweisen beim Einsatz von Vermittlungsgutscheinen. Im Teil III werden die mikroökonomischen Ergebnisse der Evaluation auf Vergleichsgruppenbasis (mit positiven Arbeitsmarktwirkungen) präsentiert. Teil IV liefert Befunde aus der makroökonomischen Wirkungsanalyse. Eine erste Kosten-Nutzen-Analyse auf noch auszubauender Datenbasis erfolgt im Teil V. Der Teil VI enthält weitere Informationen und Analysen auf der Basis von Statistiken und Befragungen zu den Empfängern und Nutzern von Vermittlungsgutscheinen sowie zur Einschaltung der privaten Arbeitsvermittler. Zur besseren Handhabbarkeit enthält Teil VIa den Text mit integrierten Tabellen, Teil VIb einen umfangreichen Tabellenanhang für vertiefende Informationen. Im Teil VII werden alle bislang vorliegenden Ergebnisse des Projekts zusammengefasst; auf dieser Basis werden Empfehlungen zur Weiterentwicklung des Vermittlungsgutscheins gegeben.

1 Einführung

Die Wirkungsweise des Vermittlungsgutscheins ist aus theoretischer Sicht relativ einfach. Die zugrundeliegende Hypothese lautet, dass private Arbeitsvermittlungen bei der Vermittlung von bestimmten Arbeitslosen effektiver arbeiten als die staatlichen Arbeitsagenturen. Dies kann z. B. durch eine Überlastung der Arbeitsagenturen zu erklären sein. Allerdings nimmt kaum ein Arbeitsloser die privaten Agenturen in Anspruch, so lange er deren Dienste selbst bezahlen muss, während ihm diejenigen der Arbeitsagentur unentgeltlich zur Verfügung stehen. Um diese Wettbewerbsverzerrung zumindest abzumildern und die höhere Effektivität der privaten Anbieter zu nutzen, übernimmt die Bundesagentur für Arbeit deren Bezahlung im Erfolgsfall durch den Vermittlungsgutschein. Sofern die privaten Vermittler wirklich in einem Teil der Fälle effektiver arbeiten und die Vermittlungsprämie einen ausreichenden Anreiz für privatwirtschaftliche Aktivität in diesem Bereich darstellt, führt dies dazu, dass teilnehmende Arbeitslose im Durchschnitt schneller eine Beschäftigung am ersten Arbeitsmarkt erhalten.

Aus theoretischer Sicht sollte somit das Vorzeichen der Effekte auf der Mikroebene eindeutig sein¹, da der Vermittlungsgutschein eine Option für den Arbeitslosen darstellt, die er nur nutzen wird, wenn seine Chancen dadurch verbessert werden. Diese Hypothese sowie die Größe der Effekte sollen im vorliegenden Forschungsbericht empirisch untersucht werden. Da dieses neuartige Instrument erst seit 2002 existiert, gibt es dazu noch keine anderen Untersuchungen.²

Deskriptive Analysen reichen nicht aus, um die Wirkung des Vermittlungsgutscheines zu beurteilen. Diese Analysen legen den Schluss nahe, dass die Gruppe der Arbeitslosen, die einen Vermittlungsgutschein erhalten haben, keine Zufallsauswahl aus der Gruppe aller Arbeitslosen ist. Es scheinen somit Selektionsmechanismen zu existieren, die auch für den Erfolg der Vermittlungsgutscheine relevant sein können. In der Mikroanalyse wird die Selektivität der Ausgabe berücksichtigt. Dadurch wird eine Ab-

¹ Zu diesem Ergebnis kommt auch Ullrich (2004) in einem suchtheoretischen Modell.

² Auch Untersuchungen zu vergleichbaren Instrumenten in anderen Ländern sind den Verfassern nicht bekannt.

schätzung der Selektionseffekte und der kausalen Effekte der Gutscheine möglich. Die methodischen Grundlagen werden sehr ausführlich in Abschnitt 2 dargelegt. Der Schwerpunkt liegt dabei auf den nicht-parametrischen Matching-Verfahren, die für die vorliegende Untersuchung verwendet werden. Die dafür erforderliche Annahme der bedingten Unabhängigkeit (CIA) wird ausführlich diskutiert. Abschnitt 3 beschreibt den verwendeten Datensatz. Zugrundegelegt wird dabei die in Teil I des Forschungsberichts beschriebene Datenstruktur, die noch leicht zu modifizieren ist. In Abschnitt 4 werden die Untersuchungsergebnisse dargestellt. Zunächst werden anhand der Mikrodaten die Determinanten der Ausgabe von Vermittlungsgutscheinen untersucht. Für die Vermittlungsgutscheine, die im Mai und Juni 2003 ausgegeben wurden, wird der Effekt auf die Wahrscheinlichkeit der Einmündung in eine reguläre Beschäftigung innerhalb von sechs Monaten abgeschätzt. Abschnitt 5 fasst die Ergebnisse zusammen.

2 Methodische Überlegungen

In diesem Abschnitt wird die Methodik beschrieben, mit der die individuellen Effekte der betrachteten Instrumente geschätzt werden sollen. Zunächst erfolgt eine Einordnung dieser Analyse, wobei auch auf die makroökonomische Untersuchung (Teil IV des Forschungsberichtes) eingegangen wird, bei der die gesamtwirtschaftlichen Effekte abgeschätzt werden (Abschnitt 2.1). Anschließend wird in Abschnitt 2.2 das mikroökonomische Evaluationsproblem dargestellt. Dieses besteht darin, dass für eine bestimmte Person, die an einer Maßnahme teilnimmt, nicht beobachtet werden kann, was ohne die Maßnahmenteilnahme passiert wäre. Der Unterschied zwischen dieser hypothetischen und der tatsächlichen Situation stellt aber gerade den Effekt der Maßnahme dar. In Abschnitt 2.3 wird erläutert, welche grundsätzlichen Ansätze zur Lösung des Evaluationsproblems zur Verfügung stehen. Zum besseren Verständnis der Problematik wird auch beschrieben, wie das Evaluationsproblem mit Hilfe eines sozialen Experimentes gelöst werden kann.

Da für die betrachteten Instrumente kein soziales Experiment durchgeführt wurde, muss auf eine nichtexperimentelle Methode zurückgegriffen werden. Die wohl am häufigsten angewendete Methode für derartige Untersuchungen ist das Matching-Verfahren, bei dem den Maßnahme-Teilnehmern ähnliche Nichtteilnehmer zugeordnet werden, um die hypothetische Situation ohne Maßnahme abzuschätzen. In Abschnitt 2.4 wird dieses

Verfahren in seiner Funktionsweise dargestellt und seine Anwendbarkeit bei unserer Untersuchung diskutiert und bejaht. In Abschnitt 2.5 werden dann verschiedene methodische Feinheiten erläutert.

2.1 Individualeffekte versus gesamtwirtschaftliche Effekte

Die vom ZEW durchgeführte mikroökonomische Evaluationsstudie dient zur Abschätzung des Individualeffektes einer Maßnahme. Es soll damit die Frage beantwortet werden, wie die Situation eines einzelnen Arbeitslosen zu einem bestimmten Zeitpunkt dadurch beeinflusst wird, dass er an einer Maßnahme teilgenommen hat. Genauer: Hat sich die Wahrscheinlichkeit des Arbeitslosen A, bis Ende des Jahres 2003 in eine reguläre Beschäftigung zu gelangen, dadurch verbessert, dass er z. B. am 1. Juni 2003 einen Vermittlungsgutschein erhalten hat?³ Dies ist definiert als der direkte Effekt für die Teilnehmer.

Falls Arbeitslose auch ohne Vermittlungsgutscheine in Beschäftigung gekommen wären, liegt aus ökonomischer Sicht ein Mitnahmeeffekt vor. Die Höhe dieser Mitnahmeeffekte kann durch die mikroökonomische Evaluationsstudie abgeschätzt werden.

Der Einsatz von Maßnahmen kann jedoch auch nicht intendierte indirekte Effekte auf andere Personen als den Teilnehmer haben. So ist es beispielsweise möglich, dass ein Unternehmen zwar den Teilnehmer einstellt, aber dafür einen anderen Arbeitslosen (Nichtteilnehmer) nicht, den es ansonsten eingestellt hätte. Dies wird häufig als Substitutionseffekt bezeichnet. Falls die Maßnahme einen Mehraufwand erzeugt, muss dies über Steuern oder Beiträge finanziert werden, was wiederum negative Effekte auf die Nichtteilnehmer (und auch auf die Teilnehmer selbst) haben kann.⁴

³ Für den einzelnen Arbeitslosen lassen sich keine Wahrscheinlichkeiten beobachten, sondern nur Zustände (also beschäftigt oder nicht). Ermittelt man jedoch die Anteile für eine homogene Gruppe von Arbeitslosen, so können diese als Wahrscheinlichkeiten interpretiert werden.

⁴ Vom Verdrängungseffekt wird gesprochen, wenn ein Unternehmen, das viele Teilnehmer einstellt, einen Vorteil am Gütermarkt erhält, der auf Kosten von anderen Unternehmen geht. Diese anderen Unternehmen müssen dann gegebenenfalls Beschäftigung abbauen. Dies dürfte beim VGS nicht so relevant sein, da Unternehmen, die Arbeitslose mit VGS einstellen keinen nennenswerten Vorteil gegenüber der Konkurrenz erzielen sollten (sofern es nicht im großen Umfang zu Missbrauch kommt).

Die Summe aller direkten und indirekten Effekte bildet den gesamtwirtschaftlichen Effekt der Maßnahme auf aggregierte Größen wie Beschäftigung oder Arbeitslosigkeit.

Da die indirekten Effekte durch eine mikroökonomische Untersuchung nicht geschätzt werden können, ist für die Abschätzung des gesamtwirtschaftlichen Effektes eine makroökonomische Untersuchung notwendig (vgl. Teil IV des Forschungsberichtes). Die mikroökonomische Untersuchung ist insbesondere wertvoll, weil dadurch die Effekte differenziert nach verschiedenen Personen-, Zeit- und regionalen Merkmalen geschätzt werden können. So ist es bspw. möglich, zu untersuchen, ob gut ausgebildete Arbeitslose stärker von der Maßnahme profitieren, ob die Maßnahme nach langer Arbeitslosigkeit geringere Wirkung zeigt, oder ob die Effekte im Arbeitsagenturbezirk A größer sind als in B. Eine solche Differenzierung ist bei einer makroökonomischen Untersuchung wegen des höheren Aggregationsniveaus nicht möglich.

2.2 Das mikroökonomische Evaluationsproblem

Der Effekt einer Maßnahme wird in der Evaluationsliteratur als treatment effect bezeichnet. Der treatment effect Δ_i auf eine Person i ist die Differenz der Ergebnisvariable Y_i im Fall mit Maßnahmenteilnahme (Y_{1i}) und im Fall ohne Maßnahmenteilnahme (Y_{0i}), also:

$$\Delta_i = Y_{1i} - Y_{0i} \quad (1)$$

Y_{1i} und Y_{0i} werden als potentielle Ergebnisse bezeichnet. Außerdem sei D_i eine Dummy-Variable, die den Wert 1 annimmt, wenn die Person tatsächlich an der Maßnahme teilnimmt, und sonst den Wert 0. Das Evaluationsproblem besteht nun darin, dass für eine bestimmte Person i nur jeweils eine Ausprägung beobachtet werden kann, nämlich für einen Maßnahmenteilnehmer ($D_i=1$) Y_{1i} und für die Nichtteilnehmer ($D_i=0$) Y_{0i} . Die Ergebnisse in den kontrafaktischen Situationen ($Y_{0i}, D_i=1$) sowie ($Y_{1i}, D_i=0$) sind nicht beobachtbar. Während individuelle kausale Effekte ($Y_{1i} - Y_{0i}$) nicht schätzbar sind, ist es mittels geeigneter Methoden durchaus möglich, in Stichproben die durchschnittlichen kausalen Effekte in der Population zu schätzen (vgl. Lechner 1998).

In mikroökonomischen Evaluationsstudien wird zumeist der durchschnittliche Effekt der Maßnahme auf die Teilnehmer geschätzt (average treatment effect on the treated, im Folgenden ATT):⁵

$$ATT = E(Y_1 - Y_0 | D = 1) = E(Y_1 | D = 1) - E(Y_0 | D = 1). \quad (2)$$

Der letzte Term in (2) beschreibt die erwartete Ausprägung der Ergebnisvariable im kontrafaktischen Zustand für die Teilnehmer, also für den Fall, dass der Teilnehmer nicht teilgenommen hätte. Dieser Term ist nicht beobachtbar. Eine Möglichkeit besteht nun darin, ihn unter Verwendung der beobachteten Ergebnisse der Nichtteilnehmer zu schätzen. Allerdings ist es nicht sinnvoll, dabei ohne weiteres alle Nichtteilnehmer heranzuziehen, da sich Teilnehmer und Nichtteilnehmer in (vom Untersuchenden) beobachtbaren (und möglicherweise auch unbeobachtbaren) Charakteristika unterscheiden. Diese werden im Allgemeinen auch die Ergebnisvariable beeinflussen, d. h.

$$E(Y_0 | D = 1) \neq E(Y_0 | D = 0). \quad (3)$$

Die durchschnittliche Ergebnisvariable der Nichtteilnehmer ist somit keine valide Schätzung für die kontrafaktische Situation der Teilnehmer. Die Differenz zwischen den beiden Seiten der Gleichung wird als Selektionsverzerrung bezeichnet. Sie resultiert daraus, dass die Teilnehmer keine zufällige Stichprobe sind. Vielmehr selektieren sie sich in die Maßnahme oder werden selektiert auf Basis von Charakteristika, die auch ihre Ergebnisvariable beeinflussen können.

Im Folgenden soll diese Selektion in Regressionsschreibweise mit einer möglichst allgemeinen *Ergebnisgleichung* dargestellt werden:⁶

$$\begin{aligned} Y_{1i} &= g_1(X_i) + U_{1i} \\ Y_{0i} &= g_0(X_i) + U_{0i}. \end{aligned} \quad (4)$$

⁵ E() stellt den Erwartungswertoperator dar. E(Y|D=1) ist der bedingte Erwartungswert von Y, wenn D=1 ist usw.

⁶ Die Darstellung orientiert sich an Heckman und Robb (1985) sowie Blundell und Costa Dias (2002).

In Gleichung (4) werden die potentiellen Ergebnisvariablen Y_1 und Y_0 mittels der Funktionen $g_1(\cdot)$ und $g_0(\cdot)$ in Abhängigkeit von den beobachteten Charakteristika X erklärt. Die Störterme U_{1i} und U_{0i} sind unabhängig zwischen den Individuen verteilt sowie unkorreliert mit den erklärenden Variablen X und es gilt $E(U_{1i}) = 0$ und $E(U_{0i}) = 0$.⁷

Im Rahmen dieser Darstellung wird zudem angenommen, dass sich die Teilnahmeentscheidung der Individuen, wie folgt parametrisieren lässt: Jedem Individuum i kann ein Index zugeordnet werden, der über eine Funktion f von erklärenden Variablen Z_i sowie von unbeobachteten Faktoren V_i abhängt

$$I_i = f(Z_i) + V_i . \quad (5)$$

Dieser Index spiegelt die Teilnahmeentscheidung wieder. Zu den beobachteten Merkmalen zählt beispielsweise die Arbeitslosendauer oder der Grund der Arbeitslosigkeit.

Eine Teilnahme liegt immer dann vor, wenn der Index größer null ist, d. h.

$$\begin{aligned} D_i &= 1 && \text{falls } I_i > 0 \\ D_i &= 0 && \text{sonst} \end{aligned}$$

Die Gleichung (5) soll als Selektionsgleichung bezeichnet werden.

Die beobachtete individuelle Ergebnisvariable ergibt sich als $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$. Zusammen mit (4) lässt sich dies umformulieren zu

$$Y_i = g_0(X_i) + \alpha_i(X_i) D_i + U_{0i} \quad (6)$$

dabei ist $\alpha_i(X_i)$ der Maßnahmeeffekt für Individuum i gegebenen dessen Charakteristika X_i (vgl. Smith und Todd 2003):

$$\alpha_i(X_i) = Y_{1i} - Y_{0i} = [g_1(X_i) - g_0(X_i)] + [U_{1i} - U_{0i}] \quad (7)$$

Die oben in Gleichung (3) für den allgemeinen Fall skizzierte Selektionsverzerrung tritt in der Regressionsgleichung (4) auf, wenn $E(U_i D_i) \neq 0$, d. h. der Teilnahme-Dummy ist nicht unabhängig von Merkmalen, die die

⁷ Eine Darstellung für den Fall einer linearen Ergebnisgleichung findet sich in Heckman und Robb (1985).

Ergebnisvariable beeinflussen. Dies kann zwei Ursachen haben (Heckman, Ichimura, Smith und Todd 1998: 1029): die *Selektion aufgrund beobachtbarer Variablen* und die *Selektion aufgrund unbeobachteter Variablen*. Der erste Fall tritt auf, wenn die unbeobachteten Faktoren U_{1i}, U_{0i} , die die Ergebnisvariable beeinflussen, mit den beobachteten Merkmalen Z_i , die die Teilnahmeentscheidung bestimmen, korrelieren. Beispielsweise könnte eine schlechte Arbeitsmoral (U_i , unbeobachtbar), die sich negativ auf die Vermittlungschancen in Beschäftigung auswirkt, mit einer Entlassung durch den letzten Arbeitgeber als Grund für die Arbeitslosigkeit (Z_i , beobachtbar) korrelieren. Wenn sich nun die Personen, die entlassen worden sind, eher keinen Vermittlungsgutschein ausstellen lassen, dann wird der Maßnahmeneffekt überschätzt, falls für diese Positivselektion nicht kontrolliert wird. Die *Selektion aufgrund beobachtbarer Variablen* tritt auch auf, wenn die gleichen Charakteristika sowohl die Teilnahme als auch die Ergebnisvariable beeinflussen. Zum Beispiel dürfte die Dauer der Arbeitslosigkeit sowohl die Vermittlungswahrscheinlichkeit als auch die Teilnahme an der Maßnahme maßgeblich bestimmen. Auch hier würde der Maßnahmeneffekt nicht konsistent geschätzt werden, wenn die Selektion nicht berücksichtigt wird. Zur zweiten Art der Selektion, der *Selektion aufgrund unbeobachtbarer Variablen*, kommt es bei der Korrelation zwischen den Störtermen U_i und V_i in den Gleichungen (5) und (6). So könnte beispielsweise die unbeobachtete Arbeitsmoral sowohl die Teilnahmeneigung (V_i) als auch die Beschäftigungswahrscheinlichkeit (U_i) negativ beeinflussen.⁸ Wenn sich nun die Individuen mit der schlechteren Arbeitsmoral eher keinen Vermittlungsgutschein ausstellen lassen, dann wird der Maßnahmeneffekt bei Nichtkontrolle für diese Positivselektion überschätzt. Die Verzerrungen infolge der *Selektion aufgrund unbeobachtbarer Variablen* und *aufgrund beobachtbarer Variablen* können gleichzeitig auftreten.

Eine einfache OLS-Regression liefert keinen unverzerrten mittleren Maßnahmeneffekt, weil die Annahme der Unkorreliertheit des Regressors D_i und des Fehlerterms U_i nicht erfüllt ist. Wie oben dargestellt, ist der Teilnahmedummy keine exogene Variable, sondern das beobachtbare Ergeb-

⁸ Es kann sich auch um unterschiedliche unbeobachtete Merkmale handeln, die Teilnahmeneigung bzw. die Beschäftigungswahrscheinlichkeit beeinflussen.

nis einer Teilnahmeentscheidung (Selektionsgleichung). Demzufolge kann eine statistische Abhängigkeit von D_i und U_i auftreten, wenn eine Selektion aufgrund beobachtbarer Variablen vorliegt, also die erklärenden Variablen Z_i aus der Selektionsgleichung (5) mit U_i korrelieren oder die erklärenden Variablen Z_i und X_i gleich sind. Die Annahme ist auch bei der Selektion aufgrund unbeobachteter Variablen verletzt (Korrelation zwischen U_i und V_i). Die Anwendung einer OLS-Regression bei einem endogenen Teilnahmedummy als erklärende Variable führt zur Über- oder Unterschätzung des Maßnahmeneffektes α . Damit ist die OLS-Regression kein geeignetes Schätzverfahren, um den Effekt einer Teilnahme auf die Ergebnisvariable zu erfassen.

2.3 Drei grundsätzliche Ansätze zur Lösung des Evaluationsproblems

Um den Maßnahmeneffekt zu schätzen, ist grundsätzlich ein Vergleich zwischen Teilnehmern und Nichtteilnehmern erforderlich, um das Ergebnis des Teilnehmers im kontrafaktischen Zustand zu konstruieren. Dabei kommen der Vorher-Nachher-Schätzer, der Querschnitts-Schätzer und der Differenz-von-Differenzen-Schätzer in Betracht (vgl. Heckman, LaLonde, Smith 1999: 1891 ff.). Diese werden im Folgenden beschrieben.

Der Vorher-Nachher-Schätzer vergleicht einen Teilnehmer mit sich selbst vor und nach der Teilnahme. Für diesen Ansatz sind somit Längsschnittdaten erforderlich. Der Vorteil dieses Ansatzes besteht darin, dass keine Daten über Nichtteilnehmer benötigt werden. Der Ansatz hat jedoch den Nachteil, dass die Methode verzerrte Schätzungen liefert, wenn es zeitvariierende unbeobachtete Effekte gibt. Darunter fallen insbesondere makroökonomische Einflüsse und Entwicklungen über den Lebenszyklus, soweit dafür nicht durch zusätzliche Regressoren (z. B. Makroindikatoren oder Alter) kontrolliert werden kann.

Die Querschnitts-Schätzer basieren auf einem Vergleich zwischen Teilnehmern und Nichtteilnehmern. Dabei muss mit geeigneten ökonometrischen Methoden für die oben erläuterte Selektionsverzerrung kontrolliert werden. Für die Analyse ist von großer Bedeutung, ob die Merkmale, die sowohl die Selektion in die Maßnahme als auch das Ergebnis beeinflussen, für den Ökonometriker beobachtbar oder unbeobachtbar sind. Durch das

in Abschnitt 2.4 beschriebene Matching-Verfahren kann nur für Unterschiede in beobachtbaren Merkmalen kontrolliert werden, nicht jedoch für Selektionsverzerrung aufgrund unbeobachteter Merkmale. Diese kann durch mikroökonomische Selektionsmodelle berücksichtigt werden. Für eine Identifizierung des Maßnahmeeffektes wird dabei eine so genannte Instrumentvariable benötigt. Dies ist eine Variable, die einerseits die Wahrscheinlichkeit der Maßnahmenteilnahme eines Arbeitslosen beeinflusst und andererseits nicht direkt mit seinem Ergebnis korreliert ist, sondern nur über die Veränderung der Teilnahmewahrscheinlichkeit. Da wir uns für den Zwischenbericht auf das Matching-Verfahren konzentrieren möchten, wird auf die mikroökonomischen Selektionsmodelle hier nicht weiter eingegangen (vgl. dazu Hagen und Steiner 2000: 65 ff.).

Als dritter grundsätzlicher Ansatz zur Lösung des Evaluationsproblems sei der Differenz-von-Differenzen-Schätzer (DvD-Schätzer) erwähnt. Dieser stellt eine Kombination aus den beiden zuvor beschriebenen Ansätzen dar. Voraussetzung ist, dass sowohl Teilnehmer als auch Nichtteilnehmer zu mindestens einem Zeitpunkt vor und einem Zeitpunkt nach der Maßnahme beobachtet werden. Der Effekt der Maßnahme für einen Teilnehmer mit einer bestimmten Ausprägung von X wird geschätzt durch:

$$\Delta^{DvD} = \left[E(Y_t | D = 1, X = x) - E(Y_t | D = 1, X = x) \right] - \left[E(Y_t | D = 0, X = x) - E(Y_t | D = 0, X = x) \right] \quad (8)$$

Für beide Gruppen wird also die durchschnittliche Differenz der Ergebnisvariable für den Zeitpunkt vor und den Zeitpunkt nach der Maßnahme gebildet. Anschließend wird die Differenz dieser Differenzen berechnet. Im Gegensatz zum Querschnitts-Schätzer führt bei diesem Schätzer Selektion aufgrund unbeobachteter Merkmale nicht zu einer Verzerrung, wenn der Einfluss dieser Merkmale auf das Ergebnis im Zeitablauf konstant ist (unbeobachtbare fixe Effekte). Die erforderlichen Annahmen sind also beim DvD-Schätzer weniger streng. Eine Verzerrung entsteht allerdings trotzdem bei Selektion aufgrund von unbeobachtbaren zeitvariierenden Effekten.

2.4 Vergleich von Teilnehmern und Nichtteilnehmern: Matching

2.4.1 Annahmen

Das Ziel der Matching-Methode ist es, die zu erwartende, nicht beobachtbare Situation der Nichtteilnahme der Teilnehmer (kontrafaktische Situation) mit einer Kontrollgruppe zu simulieren. Mit dem Matchingverfahren wird unter der Annahme, dass die Selektion in die Maßnahme nur von beobachtbaren Charakteristika abhängt, jedem Teilnehmer eine Vergleichsperson aus der Nichtteilnehmergruppe zugematcht. Die Wahl des Matches wird durch beobachtbare Merkmale bestimmt, die sich bei den gebildeten Paaren stark ähneln sollen. Der mittlere Maßnahmeneffekt für die Teilnehmer wird als durchschnittliche Differenz der Ergebnisvariablen zwischen der Teilnehmer- und Kontrollgruppe berechnet.

Das Verfahren beruht auf zwei Annahmen. Zum ersten ist das die Conditional Mean Independence Annahme, CMIA (Rosenbaum und Rubin 1983):

$$E(Y_o|D = 1, X = x) = E(Y_o|D = 0, X = x) \quad (9)$$

Diese Annahme besagt, dass der Erwartungswert der Ergebnisvariable der Teilnehmer ($D=1$), hätten diese Personen nicht an der Maßnahme teilgenommen, dem Erwartungswert in der Vergleichsgruppe der Nichtteilnehmer entspricht, wobei die beiden Gruppen hinsichtlich der beobachtbaren Charakteristika X gleich sind. Unter dieser Annahme können die Nichtteilnehmer genutzt werden, um die Ausprägung der Ergebnisvariable unverzerrt zu schätzen, welche die Teilnehmer im Durchschnitt hätten, hätten sie nicht teilgenommen. Damit kann also der letzte Term in Gleichung (2) identifiziert werden. Der Vektor X sollte dabei alle Charakteristika enthalten, die sich sowohl auf die Teilnahme an der Maßnahme als auch auf die Ergebnisvariable auswirken.

Die stärkere Conditional Independence Annahme (CIA) ist nicht nötig. Diese besagt, dass die Ergebnisvariablen unabhängig von D und X sind, während die CMIA diese Unabhängigkeit nur im Durchschnitt verlangt. Da es sich beim ATT ebenfalls nur um eine Berechnung am Mittelwert handelt, ist die CMIA ausreichend, wobei die CIA die CMIA impliziert. Mit der CMIA kann schließlich der letzte Term in Gleichung (2) identifiziert werden.

Die CMIA impliziert, dass die Teilnahmewahrscheinlichkeit nicht von unbeobachtbaren Faktoren abhängen darf, die zugleich das Ergebnis beeinflussen. Demnach ist die Annahme verletzt, wenn beispielsweise das Ergebnis im Fall ohne Maßnahme für die motivierten Personen größer ist und diese auch mit einer höheren Wahrscheinlichkeit an der Maßnahme teilnehmen, die Motivation aber durch die berücksichtigten Variablen nicht adäquat abgebildet wird. Die Nichtberücksichtigung dieser unbeobachtbaren Faktoren führt zu einer Selektionsverzerrung B (für „bias“), die entsprechend Heckman, Ichimura und Todd (1997) folgendermaßen dargestellt werden kann:

$$B(X) = E(Y_o | D = 1, X = x) - E(Y_o | D = 0, X = x) \quad (10)$$

Die zweite grundlegende Bedingung für die Validität des Matchingverfahrens ist die so genannte Stable-Unit-Treatment-Value-Assumption SUTVA (Rubin 1980). Diese besagt, dass die Ergebnisvariablen der Teilnehmer und der Nichtteilnehmer nicht dadurch beeinflusst werden, wie viele und welche anderen Individuen teilnehmen. Laut dieser Annahme darf es somit keine indirekten Effekte geben. Diese Annahme ist wichtig, wenn das Ziel der Mikroanalyse die umfassende Einschätzung der Maßnahme sein soll. Denn aus gesamtwirtschaftlicher Perspektive würde beispielsweise bei negativen indirekten Effekten der Maßnahmeeffekt überschätzt werden. Um das Instrument und seine Wirkung aus gesamtwirtschaftlicher Perspektive vollständig beurteilen zu können, ist die Berücksichtigung der indirekten Effekte in einer Makroanalyse unerlässlich.

Trotzdem haben die Ergebnisse des Matchingverfahrens auf der Mikroebene Aussagekraft, selbst wenn indirekte Effekte vorliegen. In diesem Fall kann der geschätzte Effekt mit Hilfe des folgenden gedanklichen Experimentes interpretiert werden: Aus der Gruppe der Teilnehmer wird zufällig eine einzige Person herausgezogen. Lässt man diese Person nicht an der Maßnahme teilnehmen, so entspricht die erwartete Veränderung ihrer Outcomevariable gerade dem geschätzten Effekt mit umgekehrtem Vorzeichen. Der geschätzte Effekt ist in diesem Sinne als kausal zu interpretieren. Die indirekten Effekte sind bei der Änderung des Status nur einer Person so gering, dass es keine Rückwirkungen auf die Person selbst gibt. Wechselt hingegen eine größere Zahl von Personen in die oder aus der Maßnahme, so weicht die erwartete Änderung ihrer Outcomevariable vom geschätzten Effekt aus der Mikroanalyse ab, da dann die indirekten Effekte

auch auf diese Personen selbst wirken. Für diese Politikvariation ist der geschätzte Effekt nicht mehr als kausal zu interpretieren.

Im Folgenden werden verschiedene Matching Methoden beschrieben, wobei vorweggenommen werden soll, dass man mit keiner der Methoden das Selektionsproblem mit Sicherheit lösen kann. Neuere Studien zeigen vielmehr, dass die Qualität des Matchings immer vom Zusammenspiel zwischen der gewählten Methode, den berücksichtigten Variablen, der Wahl der Stichprobe und der Datenbeschaffenheit abhängt (vgl. Heckman, Ichimura und Todd 1998: 607 und Smith und Todd 2003).

2.4.2 Drei Varianten des Matching

2.4.2.1 Exaktes Matching

Bei dieser Variante wird jedem Teilnehmer eine Person aus der Gruppe der Nichtteilnehmer zugeordnet, die exakt die gleichen beobachtbaren Merkmale aufweist. Dieses Verfahren setzt voraus, dass die Bildung einer Kontrollgruppe durch den direkten Vergleich einzelner Merkmale unter Erfüllung der CMIA auch möglich ist. Das heißt, zu jedem Teilnehmer muss es einen Nichtteilnehmer geben, der hinsichtlich aller Charakteristika die gleichen Ausprägungen aufweist (Common Support).

Der Vorteil der Methode liegt darin, dass keine Annahmen über Schätzfunktionen getroffen werden müssen, welche die Flexibilität des Verfahrens einschränken. Allerdings ist die Durchführung dieser Methode bei einer großen Anzahl von Charakteristika mit vielen Ausprägungen nicht mehr möglich. Denn es ist unwahrscheinlich, dass für jeden Teilnehmer ein gleicher Nichtteilnehmer bezüglich aller Merkmale X gefunden werden kann, wenn der Vektor X besonders groß ist. Gibt es z. B. zwanzig dichotome Merkmale, so liegt die Zahl der Kombinationsmöglichkeiten bereits über einer Million. In der Literatur wird diese Problematik als „Fluch der Dimensionalität“ bezeichnet.

2.4.2.2 Propensity Score Matching

Im Gegensatz zum exakten Matching, ist das Propensity Score Matching auch bei einer großen Anzahl von Merkmalen anwendbar. Rosenbaum und Rubin (1983) haben gezeigt, dass es ausreicht, den Teilnehmern ihre Kontrollpersonen auf der Basis der Teilnahmewahrscheinlichkeit zuzuordnen.

Wenn die Conditional Mean Independence Annahme gilt, dann gilt bei Gleichheit der Teilnehmerwahrscheinlichkeit:

$$E(Y_o|D = 1, p(X)) = E(Y_o|D = 0, p(X)), \quad (11)$$

wobei $p(X) \equiv P(D = 1|X)$ die Teilnahmewahrscheinlichkeit in Abhängigkeit von den Charakteristika ist und auch als Propensity Score bezeichnet wird. Mit Hilfe des Propensity Scores kann der mehrdimensionale Vektor X zu einer eindimensionalen Wahrscheinlichkeit $p(X)$ reduziert werden, auf deren Basis leichter eine gute Zuordnung der Nichtteilnehmer zu den Teilnehmern stattfinden kann. Der „Fluch der Dimensionalität“ kann so umgangen werden. Um den Maßnahmeneffekt zu bestimmen, benötigt man auch einen Schätzwert für den Propensity Score, da dieser nicht zu beobachten ist. Für die Schätzung der Teilnahmewahrscheinlichkeit werden flexible Logit- oder Probitmodelle verwendet, in die der Vektor X und verschiedene Funktionen von X , zum Beispiel quadratische und Interaktionsterme, eingehen.

Die Gleichung (11) kann auch allgemeiner gefasst werden:

$$E(Y_o|D = 1, b(X) = b(x)) = E(Y_o|D = 0, b(X) = b(x)), \quad (12)$$

wobei $b(x)$ der sogenannte Balancing Score (BS) ist. Der Balancing Score ist eine Funktion, welche die Teilnahmeneigung einer Person angibt. Dies muss nicht notwendigerweise die Teilnahmewahrscheinlichkeit sein; es kann sich auch um einen linearen Index⁹ oder einen Vektor handeln. Balancing bedeutet in diesem Zusammenhang, dass es bei gegebenem Wert von $b(x)$ keine systematische Abweichung zwischen der Gruppe der Teilnehmer und der Nichtteilnehmergruppe hinsichtlich des Ergebnisses im Fall der Nichtteilnahme gibt. Ein Balancing Score muss eine umfassendere Informationsmenge sein als der Propensity Score, d. h. wenn $b(x_1) = b(x_2)$ gilt, dann gilt auch $p(x_1) = p(x_2)$, aber nicht umgekehrt (zum Beweis vgl. Rosenbaum und Rubin 1983). In der Praxis ist der Propensity Score als Index für die Teilnahmeneigung am weitesten verbreitet. Die Verwendung eines Balancing Scores ist jedoch erforderlich, wenn der ATT für eine Subpopulation geschätzt werden soll, die durch eine Kompo-

⁹ Z. B. kann als Teilnahmeneigung ein linearer Index aus den Koeffizienten aus einem Wahrscheinlichkeitsmodell und den Charakteristika verwendet werden

nente von X definiert wird (z. B. Geschlecht). Es wird dann nicht nur auf den Propensity Score, sondern zusätzlich auf diese Variable gematcht.

Für das Propensity Score Matching ist eine weitere Annahme nötig, nämlich die des Common Support. Um den ATT schätzen zu können, muss demnach die prognostizierte Wahrscheinlichkeit der Teilnahme an einer Maßnahme strikt kleiner eins sein,

$$P(D = 1|X) < 1, \quad (13)$$

für alle Ausprägungen der Variablen X , die in der Evaluation berücksichtigt werden.¹⁰ Die Bedingung stellt sicher, dass es in der Stichprobe für jeden beobachteten X -Wert sowohl Teilnehmer als auch Nichtteilnehmer an einer Maßnahme gibt. Heckman, Ichimura und Todd (1997) weisen daraufhin, dass der nichtgarantierte Common Support für beide Personengruppen ein großes Problem bei den nicht-experimentellen Methoden ist, das bei den experimentellen Methoden nicht auftritt, weil dafür explizit kontrolliert wird. Beispielsweise kann kein Maßnahmeneffekt für über 50-Jährige berechnet werden, wenn für diese Personengruppe keine Nichtteilnehmer vorhanden sind.

Wenn die CMIA-Annahme erfüllt ist, dann kann das Propensity Score Matching folgendermaßen durchgeführt werden: Im ersten Schritt wird mit Hilfe eines Wahrscheinlichkeitsmodells die Teilnahme durch den Vektor X erklärt. Zur Schätzung werden dabei in der Regel parametrische Ansätze, wie z. B. das Logit- und Probitmodell benutzt. Danach wird für jeden Teilnehmer und Nichtteilnehmer der Propensity Score basierend auf der Schätzung vorausgesagt. Nun werden für jeden Teilnehmer i vergleichbare Nichtteilnehmer j gesucht, wobei ein Maß benutzt wird, das die Ähnlichkeit der Propensity Score Werte der Teilnehmer und Nichtteilnehmer quantifiziert, $|p_i - p_j|$.

Verschiedene Versionen des Propensity Score Matchings

Der Maßnahmeneffekt für den Teilnehmer i wird als Differenz aus der Ergebnisvariable des Teilnehmers i und dem gewichteten Durchschnitt der

¹⁰ Die strengere Annahme $0 < P(D=1|X) < 1$ ist nur erforderlich, wenn auch für die Nichtteilnehmer der Effekt einer hypothetischen Teilnahme geschätzt werden soll.

Ergebnisvariablen der zugeordneten Kontrollpersonen errechnet. Damit wird der ATT wie folgt bestimmt:

$$\frac{1}{N_1} \sum_{i=1}^{N_1} \left(Y_{1i} - \sum_{j=1}^{N_0} w(i, j) Y_{0j} \right), \quad (14)$$

wobei N_1 die Anzahl der Teilnehmer und N_0 die Anzahl der Nichtteilnehmer angibt.

In der mikroökonomischen Evaluationsforschung gibt es verschiedene Matching Ansätze, die sich in den Gewichten $w(i, j) \in [0,1]$ unterscheiden, mit denen die Nichtteilnehmer in die Berechnung eingehen. Die verschiedenen Ansätze werden in Heckman, Ichimura, Smith and Todd (1998) ausführlich diskutiert.

Beim traditionellen paarweisen Matching, das auch Single Nearest-Neighbor Matching (NN-Matching) genannt wird, wird jedem Teilnehmer i nur die Kontrollperson j aus der Gruppe der Nichtteilnehmer zugeordnet, die zur kleinsten Distanz $|p_i - p_j|$ führt. Dieser nächste Nachbar erhält als Gewicht $w(i, j) = 1$, sodass bei diesem Ansatz nur eine Kontrollperson für jeden Teilnehmer berücksichtigt wird. Man kann beim NN-Matching zwischen dem Matching mit und ohne Zurücklegen der bereits verwendeten Kontrollpersonen unterscheiden. Im ersten Fall können die Nichtteilnehmer mehrmals als Kontrollperson fungieren, während im zweiten Fall jede Kontrollperson nur einmal genutzt wird. Zwischen beiden Varianten gibt es einen Trade-off zwischen der Verzerrung und der Varianz des geschätzten Effektes. Beim NN-Matching mit Zurücklegen steigt zwar die Varianz¹¹ des Schätzers, aber auch die Qualität des Matches. Die Zuordnung findet nämlich nur auf der Basis der Propensity Score Ähnlichkeit statt, ohne Rücksicht auf eine mögliche vorherige Verwendung des Nichtteilnehmers als Kontrollperson. Beim Matching ohne Zurücklegen könnte es beispielsweise vorkommen, dass dem Teilnehmer i die Kontrollperson j am ähnlichsten ist, diese aber bereits verwendet wurde. Dann muss ein anderer Nichtteilnehmer herangezogen werden, dessen Propensity Score weiter entfernt ist

¹¹ Die Varianz ist die mittlere quadratische Abweichung zwischen dem geschätzten Effekt und dem tatsächlichen Effekt. Im Allgemeinen gilt, je kleiner die Varianz ist, desto effizienter ist der Schätzer.

als der von j . Hinzu kommt, dass der geschätzte Effekt und die Qualität des Matchings von der Reihenfolge der Zuordnung abhängen. Beim NN-Matching ohne Zurücklegen nimmt die Qualität des Matchings umso mehr ab, je weniger Nichtteilnehmer sich im Datensatz befinden. Wenn die Zahl der Nichtteilnehmer kleiner ist als die der Teilnehmer, ist die Methode nicht durchführbar. Da die Datensätze in der Praxis oft klein sind, wird in Evaluationsstudien häufig die Methode mit Zurücklegen angewendet. Da bei unserer Untersuchung ein sehr großer Datensatz zugrunde liegt und die Teilnehmer nur einen kleinen einstelligen Prozentsatz der Beobachtungen ausmachen, sollte dieser Aspekt von geringer Bedeutung sein. In der Tat unterscheiden sich die Ergebnisse bei den beiden Varianten nur marginal (vgl. Abschnitt 4.2).

Eine modifizierte Version vom Single NN-Matching ist das so genannte Caliper Matching (Cochran and Rubin, 1973). Hier wird ein Toleranzbereich definiert, in dem die Distanz $|p_i - p_j|$ liegen muss, damit der nächste Nachbar noch als Kontrollperson akzeptiert wird. Wenn für einen Teilnehmer i kein nächster Nachbar innerhalb des Bereichs vorliegt, dann wird diese Person in der Schätzung des kausalen Effektes nicht berücksichtigt. Mit dieser Methode wird die Wahrscheinlichkeit eines schlechten Matches zwar reduziert, aber es ist im Vorfeld nicht klar, wie groß der Toleranzbereich sein sollte.

Neben dem Single NN-Matching gibt es auch eine NN-Variante, bei der jedem Teilnehmer mehrere (z. B. zehn) Nichtteilnehmer zugeordnet werden, die alle das gleiche Gewicht haben. So werden nicht nur die Informationen eines Nichtteilnehmers, sondern die von mehreren genutzt.

Beim so genannten Kernel Matching werden ebenfalls die Informationen mehrerer Kontrollpersonen, im Extremfall aller Nichtteilnehmer, berücksichtigt. Ein Kernel ist dabei eine Funktion, die den Beitrag der Ergebnisvariablen jeder Kontrollperson in der Berechnung des kausalen Effektes gewichtet (vgl. (15)). Im allgemeinen wird so gewichtet, dass die Kontrollpersonen, die dem Teilnehmer i ähnlicher sind, ein höheres Gewicht erhalten, als solche, deren Propensity Score weiter entfernt ist.

Das Gewicht wird beim Kernel Matching wie folgt bestimmt:

$$w(i, j) = \frac{K_{ij}}{\sum_{j=1}^{N_0} K_{ij}} \quad (15)$$

wobei $K_{ij}=K((p_i-p_j)/h)$ die Kernelfunktion ist, die entfernte Propensity Scores der Nichtteilnehmer geringer gewichtet. h ist die so genannte Bandweite. Am häufigsten wird die Standardnormalverteilung (mit Mittelwert 0) als Kernel verwendet:

$$K(\varphi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\varphi^2\right) \text{ mit } \varphi = \frac{p_i - p_j}{h} \quad (16)$$

Härdle (1990) hat herausgestellt, dass die Wahl der Bandweite h und nicht die Wahl der Kernelfunktion für die nicht-parametrische Anpassung, also die Qualität des Matches entscheidend ist. Die Bandweite gibt an, mit welcher Geschwindigkeit die Gewichte bei zunehmender Distanz $|p_i - p_j|$ sinken. Für die Bestimmung der Bandweite gibt es keine goldene Regel. Es gibt einen Trade-off zwischen der Varianz und der Selektionsverzerrung des Schätzers. Bei großem h werden viele Beobachtungen in der Schätzung berücksichtigt, was zu einer sinkenden Varianz aber einer steigenden Selektionsverzerrung führt, da die zusätzlich berücksichtigten Nichtteilnehmer weiter vom betrachteten Teilnehmer entfernt sind. Da beim Matching für die Selektion in die Teilnehmer- und Nichtteilnehmergruppe auf der Basis von beobachtbaren Charakteristika kontrolliert werden soll, sollte die Bandweite kleiner sein. Eine Möglichkeit, die bei der Bestimmung der Bandweite sehr häufig genutzt wird, ist die „Silverman Rule of Thumb“ (ROT):

$$h_{ROT} = 0.9A \cdot n^{-1/5} \quad (17)$$

Wobei h die gewählte Bandweite ist und $A=\min(\text{std}, \text{iqr}/1.34)$. std steht für die Standardabweichung, iqr für die Differenz zwischen dem 75 %igen Quantil und dem 25 %igen Quantil der Stichprobe und n ist der Stichprobenumfang.

Allgemein gilt, dass im Vergleich zum Single NN-Matching beim Kernel Matching die Varianz der Schätzung reduziert wird, sich dafür aber die Selektionsverzerrung möglicherweise erhöht, weil auch sehr „unähnliche“ Nichtteilnehmer genutzt werden, wenn auch mit einem geringen Gewicht.

2.4.2.3 Mahalanobis Matching

Das Mahalanobis Matching ist ein weiteres Matching Verfahren, das von Cochran und Rubin (1973) ausführlich beschrieben wird. Bei diesem Verfahren wird nicht nur wie beim Propensity Matching die Distanz zwischen den Propensity Scores der Teilnehmer und Nichtteilnehmer betrachtet. Um die Kontrollpersonen zu bestimmen, werden beim Mahalanobis Matching die Differenzen in den einzelnen Variablen mit der Kovarianzmatrix der Variablen der Nichtteilnehmer gewichtet. Die Mahalanobis Distanz wird definiert als:

$$d(X_i, X_j) = (X_i - X_j)^T Cov_j^{-1} (X_i - X_j) \quad (18)$$

wobei X_i und X_j die Vektoren der Merkmale für die Teilnehmer i und Nichtteilnehmer j sind. Cov_j ist die Kovarianzmatrix von X_j in der Stichprobe der Nichtteilnehmer. Jedem Teilnehmer wird dann der Nichtteilnehmer mit der geringsten Mahalanobis Distanz zugeordnet. Dabei können die Abstände bei Merkmalsausprägungen, die eine hohe Varianz bei den Nichtteilnehmern aufweisen, durch die Gewichtung höher ausfallen als bei Variablen mit geringer Varianz. In dieser Berücksichtigung der unterschiedlichen Schwankungen der Merkmale liegt der Vorteil dieser Methode. Wenn alle Charakteristika eine ähnliche Varianz bei den Nichtteilnehmern aufweisen, gibt es gegenüber dem PS-Matching keinen Vorteil.

Beim Mahalanobis Matching gibt es wiederum verschiedene Varianten. Das oben definierte Distanzmaß gehört zur ursprünglichen Version dieses Matching Verfahrens. Rosenbaum und Rubin (1985) haben in einer weiteren Variante sowohl die Unterschiede in den Propensity Scores als auch in den Merkmalsausprägungen zwischen dem Teilnehmer i und dem Nichtteilnehmer j berücksichtigt. Die Mahalanobis Distanz hat dann die Form:

$$d(V_i, V_j) = (V_i - V_j)^T Cov_j^{-1} (V_i - V_j) \quad (19)$$

wobei V Werte aus $\{X, p(X)\}$ beinhaltet und Cov_j die Kovarianzmatrix von $\{X, p(X)\}$ der Nichtteilnehmer darstellt.

2.4.3 Kritische Reflexion des Matching-Ansatzes

In jüngerer Vergangenheit ist in der Literatur viel darüber diskutiert worden, ob in empirischen Anwendungen das Evaluationsproblem mit Hilfe des Propensity Score (PS) Matchings tatsächlich gelöst werden kann. Lehr-

reich ist in diesem Zusammenhang die Diskussion zwischen Dehejia und Wahba auf der einen Seite sowie Smith und Todd auf der anderen Seite.

Dehejia und Wahba (1999, 2002) nutzen Propensity Score Matching Methoden, um den ATT der National Supported Work Demonstration zu schätzen, eines US-amerikanischen Arbeitsmarktprogrammes, das Problemgruppen in Trainingsmaßnahmen schickt. Sie nutzen dabei die Ergebnisse eines sozialen Experimentes (vgl. Abschnitt 2.3.2) als Benchmark. Datenbasis ist ein Sample, das LaLonde (1986) bereits für eine Untersuchung der Effektivität von nichtexperimentellen Methoden benutzte. Dehejia und Wahba (1999, 2002) verwenden von den Teilnehmern aus dem ursprünglichen LaLonde-Sample nur diejenigen, für die das Einkommen in beiden Jahren vor der Maßnahme beobachtet werden kann (DW-Sample). Die Autoren kommen zu dem Ergebnis, dass mit Matching Schätzungen des ATT möglich sind, die nah bei der Benchmark liegen und robust gegenüber der Spezifikation der Propensity Score Schätzgleichung sind. Außerdem verwenden sie einen Algorithmus, mit dessen Hilfe diese Schätzgleichung systematisch so spezifiziert wird, dass die Kovariaten in Teilnehmer- und Kontrollgruppe ausbalanciert werden (vgl. dazu Abschnitt 2.5.2). Dies sei für die Qualität der Ergebnisse sehr wichtig. Weiterhin weisen die Autoren darauf hin, dass mit Hilfe des geschätzten Propensity Score sehr gut mögliche Unterschiede zwischen Teilnehmer- und Kontrollgruppe dargestellt werden können (fehlender Common Support), was die mangelnde Vergleichbarkeit evident macht.

Smith und Todd (2004) analysieren die Effektivität des Propensity Score Matchings mit denselben Daten und kommen zu einem anderen Ergebnis. Sie können zwar die geringen Verzerrungen des geschätzten ATT für das DW-Sample reproduzieren. Allerdings erhalten sie bei Anwendung des Propensity Score Matchings auf das LaLonde-Sample sowie auf ein anderes Subsample, das ihnen sinnvoller erscheint (ST-Sample), erheblich größere Verzerrungen gegenüber der experimentellen Benchmark. Daraus ziehen sie die Schlussfolgerung, dass Dehejia und Wahba durch die Wahl ihres Subsamples das Selektionsproblem derart vereinfacht haben, dass teilweise einfache Regressionsmodelle ebenfalls zu Schätzungen des ATT führen, die nur geringfügig verzerrt sind. Die Ergebnisse jener Autoren hinsichtlich der Effektivität des Propensity Score Matchings könnten somit nicht verallgemeinert werden. Weiterhin kommen Smith und Todd (2004)

zu dem Ergebnis, dass Differenz-von-Differenzen Matching Schätzer¹² zu geringeren Verzerrungen des geschätzten ATT führen.

In seiner Antwort kritisiert Dehejia (2003), dass Smith und Todd (2004) bei der Anwendung des PS Matchings auf das LaLonde- und das ST-Sample die Schätzgleichung für den Propensity Score nicht geeignet spezifiziert hätten. Er weist darauf hin, dass für jede Kombination von Teilnehmer- und Kontrollgruppe eine eigene Spezifikation der Schätzgleichung gewählt werden muss, durch welche die Kovariaten ausbalanciert werden. Weiterhin weist er darauf hin, dass die Sensitivität der Schätzung gegenüber geringfügigen Änderungen der Spezifikation zu prüfen ist. Die Schätzungen für das LaLonde-Sample (und in geringerem Ausmaß für das ST-Sample) seien diesbezüglich sensitiv, so dass Propensity Score Matching für diese Samples nicht geeignet sei.

Schließlich führen Smith und Todd (2003) weitere Sensitivitätsanalysen durch. Diese basieren auf den folgenden Überlegungen: Für eine konsistente Schätzung des Propensity Scores kann entweder die experimentelle Teilnehmergruppe oder die experimentelle Kontrollgruppe verwendet werden, die ebenfalls teilgenommen hätte, wenn sie nicht herausgelost worden wäre. Es können auch beide Gruppen gemeinsam verwendet werden. Außerdem kann für eine konsistente Schätzung der Verzerrung, die aus dem Matching resultiert, einerseits der ATT nichtexperimentell unter Verwendung der experimentellen Teilnehmergruppe geschätzt werden, wovon dann der im sozialen Experiment ermittelte „tatsächliche“ Effekt subtrahiert wird. Andererseits kann aber die Verzerrung auch direkt geschätzt werden, indem der experimentellen Kontrollgruppe Nichtteilnehmer zugematcht werden. Schließlich ist es möglich, dass mehrere Nichtteilnehmer denselben geschätzten Propensity Score aufweisen. Falls es zu einem Teilnehmer mehrere Nichtteilnehmer mit gleicher Distanz gibt, wird beim Single Nearest Neighbor Matching zufällig entschieden, welcher Nichtteilnehmer zugematcht wird. Bei kleinen Samples kann die Schätzung dadurch

¹² Bei dieser Methode wird nach dem Matching jeweils für die Teilnehmergruppe und die gematchte Kontrollgruppe die Differenz der Zielvariable für einen Zeitpunkt vor und einen Zeitpunkt nach dem Treatment gebildet. Die Differenz dieser beiden Differenzen wird als kausaler Effekt der Maßnahme interpretiert. Durch die zusätzliche Differenzbildung kann eine verbleibende Selektionsverzerrung eliminiert werden, soweit sie einen im Zeitablauf konstanten Einfluss auf die Zielvariable hat.

einem deutlichen Zufallseinfluss unterliegen. Die Autoren finden heraus, dass die Schätzergebnisse sämtlicher vorhergehender Analysen eine hohe Sensitivität gegenüber diesen drei Fragestellungen aufweisen. Dies ist dadurch zu erklären, dass sowohl das experimentelle Sample als auch die Zahl der verwendbaren nichtexperimentellen Nichtteilnehmer klein ist, was sich auch in hohen Standardfehlern der Matching Schätzer widerspiegelt. Die Autoren ziehen die Schlussfolgerung, dass die vorherigen Ergebnisse aus diesem Grunde vorsichtig zu interpretieren sind.

Für die vorliegende Untersuchung können aus der Diskussion die folgenden Schlussfolgerungen gezogen werden. Selbst wenn alle Daten vorliegen, die für die Erfüllung der CMIA erforderlich sind (vgl. folgender Abschnitt), ist dadurch nicht automatisch eine unverzerrte Schätzung gewährleistet. Vielmehr muss beim Propensity Score Matching die Spezifikation der Schätzgleichung sorgfältig erfolgen, damit die beobachteten Kovariaten auch tatsächlich ausbalanciert werden. Um dies zu erreichen, stehen verschiedene Balancing Tests zur Verfügung, die in Abschnitt 2.5.2 diskutiert werden. Wenn das Balancing gelingt, ist noch anhand von Sensitivitätsanalysen zu überprüfen, ob die Schätzergebnisse robust sind. Die Beobachtungszahl sollte dabei in der vorliegenden Untersuchung kein Problem darstellen.

2.4.4 Diskussion der Matching-Annahmen für vorliegende Untersuchung

Die CMIA verlangt von uns alle Charakteristika X zu beobachten, die sowohl die Entscheidung zur Maßnahmenteilnahme als auch die Wahrscheinlichkeit der Beschäftigung beeinflussen. Die Annahme besagt, dass bedingt auf die Beobachtbarkeit aller relevanten Informationen die Teilnahme nicht mit einem späteren Arbeitsmarktstatus korreliert, den der Teilnehmer erfahren hätte, wenn er keinen Gutschein besessen hätte.

Wie plausibel ist die CMIA in Relation zur Reichhaltigkeit der verfügbaren Datensätze? Zusätzlich zu den BewA-Dateien wurden für die Evaluierung der Vermittlungsgutscheine auch die Beschäftigten-Historik, die Leistungsempfänger-Historik und die Maßnahme-Teilnahme-Gesamtdatenbank vom IAB zur Verfügung gestellt. Das ermöglicht neben der Erfassung der aktuellen Charakteristika der Arbeitslosen, wie Alter, Geschlecht, Arbeitslosig-

keitsdauer, angestrebte Berufsgruppe, auch die detaillierte Berücksichtigung der Erwerbsbiographie.

Aus Studien zur Evaluierung aktiver Arbeitsmarktpolitik (u. a. Sianesi 2003) geht hervor, dass die Entscheidung eines Arbeitslosen, an einem Programm teilzunehmen, stark von seiner subjektiven Einschätzung der eigenen Beschäftigungsaussichten abhängt. Wenn diese Wahrnehmung nah genug an der Realität ist, dann spiegelt sich darin auch die tatsächliche Arbeitsmarktchance wider. Aus der Implementationsstudie zum Projekt ist zudem bekannt, dass im Durchschnitt die Wahrscheinlichkeit, einen Gutschein zu erhalten, dann höher ist, wenn der Arbeitsvermittler gute Beschäftigungschancen für den Arbeitslosen sieht. Demnach ist es für die Analyse entscheidend, genug Informationen zu berücksichtigen, die geeignet sind, um die persönliche Einschätzung der Beschäftigungswahrscheinlichkeit sowie die Einschätzung durch den Arbeitsvermittler abzubilden.

Aus der Erwerbsbiographie:

Es gehen in die Untersuchung die Arbeitslosigkeits- und Beschäftigungsperioden von 1996 bis zum aktuellen Rand, sowie die Teilnahme an ABM und an Trainingsmaßnahmen von 2000 beginnend ein. Dabei wird die zeitliche Struktur dieser Ereignisse durch unterschiedliche Zeitintervalle berücksichtigt. Die Intervalle der näheren Vergangenheit sind kürzer, um sie stärker zu gewichten. Das heißt, für die letzten 6 Monate werden die Tage in den Zuständen (Arbeitslosigkeit, Beschäftigung, Teilnahme an ABM bzw. SAM, Teilnahme an FbW, sonstiges) monatsweise erfasst. Danach werden die Betrachtungszeiträume größer. So werden die Tage in den erwähnten Arbeitsmarktzuständen nach 12 Monaten nur noch jahresweise aufgenommen. Wir vermuten, dass die jüngere Vergangenheit wichtiger für die Teilnahme ist als die fernere und bilden sie daher differenzierter ab. Auch von einer Teilnahme an ABM bzw. SAM oder auch FbW kann bei rationalen Individuen auf eine höhere subjektive Beschäftigungswahrscheinlichkeit geschlossen werden. Die Teilnahme zeugt von dem Willen und der Bereitschaft auf den ersten Arbeitsmarkt zurückzukehren, die ohne berechtigte Hoffnung auf Beschäftigung kaum bestehen dürfte. Dabei trägt die unterschiedliche Gewichtung der Arbeitsmarktzustände, je nachdem wie weit sie zurückliegen, einer Veränderung der subjektiven Einschätzung der Beschäftigungschancen Rechnung. Zudem werden mögliche Sperrzeiten und

Krankmeldungen aus den letzten 8 Jahren berücksichtigt. Diese Variablen erfassen eine gewollte oder ungewollte mangelnde Verfügbarkeit, die die individuellen Beschäftigungschancen widerspiegeln können.

Aus den aktuellen Merkmalen des Arbeitslosen:

Zunächst einmal sind demographische Charakteristika wie Alter, Geschlecht, Nationalität, Familienstand und der ausgeübte Beruf aus dem BewA-Datensatz wichtige Determinanten der persönlichen Einschätzung der Arbeitsmarktchancen.

Auch Informationen zum Humankapital dürften besonders gute Indikatoren für die subjektiv empfundenen Beschäftigungsaussichten sein. Arbeitslose, die Qualifikationen besitzen, die am Arbeitsmarkt nicht nachgefragt werden, werden ihre Beschäftigungsaussichten eher schlecht einschätzen. Die zur Verfügung gestellten Datensätze ermöglichen eine relativ detaillierte Berücksichtigung sowohl der Schul- wie auch der Berufsausbildung. Zusätzlich gibt es Angaben der Mitarbeiter der Arbeitsämter zur Einschätzung des beruflichen Qualifikationsniveaus. Diese Angaben weichen oftmals von dem formalen Berufsausbildungsabschluss ab und spiegeln somit Qualifikationen wider, die die Individuen in ihrer Berufstätigkeit erlangt haben.

Die subjektiv wahrgenommene Beschäftigungswahrscheinlichkeit hängt auch von der Flexibilität des Arbeitslosen ab. Beispielsweise gestaltet sich die Beschäftigungsaufnahme für Frauen mit Kleinkindern und Kindern im schulpflichtigen Alter aufgrund weniger und inflexibler institutioneller Betreuungsmöglichkeiten schwierig. Wir kontrollieren in unserer Analyse nicht nur für Kinder in unterschiedlichen Alterskategorien, sondern berücksichtigen zudem, ob der Arbeitslose allein erziehend ist oder in einer Partnerschaft lebt. Der Einsatz auf dem Arbeitsmarkt wird auch durch eine Behinderung eingeschränkt. Außerdem verringern sich die Beschäftigungsmöglichkeiten ebenfalls durch bestimmte Arbeitszeitwünsche von Seiten der Arbeitslosen. Die besten Arbeitsmarktaussichten wird sich der Arbeitslose ausmalen, der bereit ist, Teil- oder Vollzeit oder sogar im Schichtdienst tätig zu sein. In die gleiche Richtung geht auch die Argumentation bei Jobwechslern. Arbeitslose, die den Beruf wechseln möchten, also noch keine Arbeitserfahrungen in dem gewünschten Job haben, dürften sich bei

einer angespannten Arbeitsmarktlage eher geringe Erfolgsaussichten ausrechnen.

Zusätzliche Hinweise für die Beschäftigungswahrscheinlichkeit des Arbeitslosen kann auch der Leistungsbezug geben. Dieser ist in der Regel umso höher, je näher die letzte Beschäftigung liegt bzw. je höher der letzte Arbeitsverdienst war, wobei dieser Zusammenhang nicht linear ist. Für die beiden Größen Dauer der Arbeitslosigkeit und Höhe des letzten Arbeitsverdienstes wird kontrolliert. Ein relativ dazu hoher oder niedriger Leistungsbezug wirkt sich auf die Motivation des Arbeitslosen bei der Suche und damit auf die Einschätzung der Beschäftigungschancen aus. Für die Gültigkeit der CMIA in der vorliegenden Analyse können auch Nachfrageaspekte des Arbeitsmarktes wichtig sein. Dies soll durch die regionale Arbeitslosenquote und andere regionale Indikatoren berücksichtigt werden. Zusätzlich geht auch die zuständige Agentur ein. Wie die IAW-Analyse zur Typisierung der Agenturen gezeigt hat, gibt es Unterschiede zwischen den Agenturen im Umgang mit den Vermittlungsgutscheinen und dem Informationsfluss.

Zusammenfassend ist die CMIA aufgrund der Berücksichtigung der Erwerbsbiographie des Arbeitslosen und der Merkmale, die seine aktuelle Lage beschreiben, für die vorliegende Untersuchung vertretbar. Die wahrscheinlich verbleibenden Verzerrungen dürften hinreichend klein sein. Eine Garantie dafür kann es jedoch nicht geben. Bisherige Erfahrungen mit mikroökonomischen Analysen basierend auf den Individualdaten der BA liegen mit den vom IAB geförderten ZEW-Untersuchungen zur Wirksamkeit der gemeinnützigen Arbeitnehmerüberlassung in Rheinland-Pfalz vor, siehe Almus et al. (1998, 1999) sowie Lechner et al. (2001). Diese Untersuchungen zeigen die Vielfalt und Qualität der BA-Individualdaten auf, deren Potenzial vollständig in mikroökonomischen Analysen erschließbar wird.

Für die Anwendung des Matching-Verfahrens ist neben der CMIA noch die Annahme der nichtdeterministischen Teilnahmewahrscheinlichkeit ($P(D=1|X) < 1$) zu rechtfertigen. Diese Annahme besagt, dass eine so genannte Instrumentvariable existieren muss, die die Teilnahmeentscheidung, aber nicht die Beschäftigungswahrscheinlichkeit bestimmt. Dabei muss die Instrumentvariable nicht beobachtbar sein, von ihr muss nur ei-

ne zusätzliche Variation zwischen Teilnehmern und Nichtteilnehmern ausgehen. Das heißt, der Teilnahmestatus ist bedingt auf die Kovariate nicht vollständig vorherbestimmt, sondern zufällig. Im Fall der Vermittlungsgutscheine wird insbesondere die Ermessungsentscheidung der Sachbearbeiter in den einzelnen Arbeitsämtern die erforderliche Zufälligkeit in die Teilnahmeentscheidung sichern. Die Sachbearbeiter haben Einfluss auf die Vergabe der Vermittlungsgutscheine, aber nicht auf die Beschäftigungswahrscheinlichkeit der Arbeitslosen. Zum Beispiel könnte nur einer von zwei hinsichtlich ihrer relevanten Charakteristika, wie Alter, Bildung und Beruf, gleichen Arbeitslosen einen VGS erhalten. Die Entscheidung zur Ausgabe des Vermittlungsgutscheines liegt zum Teil im Ermessen des Sachbearbeiters und wird nicht bei Vorliegen bestimmter beobachtbarer Größen automatisch getroffen. Damit sorgen die Sachbearbeiter im Arbeitsamt für zusätzliche Variation bei der Programmteilnahme, die notwendig für die Anwendbarkeit des Matching-Verfahrens ist, aber vom Ökonometriker nicht direkt beobachtet werden muss.

2.5 Methodische Details

2.5.1 Datum des Maßnahmebeginns und zeitabhängige Variablen

Sowohl die Ergebnisvariablen als auch ein Teil der Kovariaten für das Matching können nur sinnvoll in Relation zum Datum des Maßnahmebeginns definiert werden. In die erste Kategorie fällt z. B. die Größe des Maßnahmeeffektes auf die Eingliederung in reguläre Beschäftigung 4 Monate nach Ausgabe eines Vermittlungsgutscheines, in die zweite die Dauer der Arbeitslosigkeit bis zum Maßnahmebeginn sowie der Gesundheitszustand kurz vor Maßnahmebeginn. Dieses Datum ist jedoch bei den Nichtteilnehmern nicht beobachtbar. Daher ist zu überlegen, wie für die Nichtteilnehmer ein Datum für den hypothetischen Maßnahmebeginn zu generieren ist.

Es wäre nun möglich, allen Nichtteilnehmern einen einheitlichen Zeitpunkt zuzuweisen, bspw. in der Mitte des Zeitraums, für den die Maßnahmen analysiert werden. Aus einer solchen Zuweisung würde jedoch ein methodisches Problem entstehen. Eine Festlegung des Ausgabezeitpunktes würde dazu führen, dass Variablen, die auf diesem Zeitpunkt basieren (z. B. die bisherige Dauer der Arbeitslosigkeit) nicht mehr unabhängig von der

Teilnahme an der Maßnahme VGS sind. Denn die Ausgabezeitpunkte wären für alle Nichtteilnehmer gleich, für die Teilnehmer hingegen unterschiedlich (vgl. Lechner 1999).

In der Literatur werden verschiedene Methoden vorgeschlagen, um mit diesem Problem umzugehen. Diese sollen im Folgenden dargestellt und hinsichtlich ihrer Eignung für die Zwecke des Projektes verglichen werden.

a) Aufteilung des Propensity Scores

Eine von drei Methoden, die Lechner (1999) vorschlägt, basiert darauf, den Balancing Score in zwei Komponenten aufzuteilen, von denen die eine die zeitabhängigen Variablen enthält und die andere die zeitunabhängigen. Zunächst wird mittels eines Probitmodells die Komponente geschätzt, die auf den zeitunabhängigen Variablen basiert, welche auch für die Nichtteilnehmer wohldefiniert sind. Außerdem wird (mit Hilfe der Delta-Methode) die dazugehörige Varianz abgeleitet.

Beim eigentlichen Single-Nearest-Neighbor-Matching ohne Zurücklegen wird nun wie folgt vorgegangen: für den ersten (zufällig ausgewählten) Teilnehmer wird ein 90 %-Konfidenzintervall um den partiellen Balancing Score gebildet. Falls nur einer oder keiner der Nichtteilnehmer mit seinem partiellen Balancing Score innerhalb dieses Intervalls liegt, wird der Nichtteilnehmer dem Teilnehmer zugematcht, dessen Balancing Score am nächsten liegt. Gibt es hingegen mehrere Nichtteilnehmer in dem Intervall, so werden für diese die zeitabhängigen Variablen auf Basis des Anfangsdatums des betrachteten Teilnehmers ermittelt. Anschließend wird aus dem partiellen Balancing Score und den zeitabhängigen Variablen die Mahalanobis-Distanz zwischen dem Teilnehmer und diesen Nichtteilnehmern berechnet.¹³ Der Nichtteilnehmer mit der geringsten Distanz wird dem Teilnehmer zugematcht. Für die übrigen Teilnehmer wird ebenso verfahren.

Damit mit dieser Methode der ATT konsistent geschätzt wird, sind neben der CMIA zusätzliche Annahmen erforderlich. Erstens setzt sich der Propensity Score additiv aus den genannten Komponenten zusammen (dies

¹³ Dabei wird die Kovarianz-Matrix für ein bestimmtes Datum in der Mitte des betrachteten Zeitraumes verwendet.

ist weniger restriktiv). Zweitens muss die Komponente, die auf den zeitabhängigen Variablen basiert, unabhängig von der anderen Komponente und normalverteilt sein. Lechner (1999) testet bei seiner Anwendung mit Spezifikationstests bspw. auf bedingte Homoskedastizität und Normalverteilung und kann diese nicht ablehnen.

b) Zufällige Generierung von Anfangsdaten für die Nichtteilnehmer

Bei der zweiten von Lechner (1999) vorgeschlagenen Methode wird jedem Nichtteilnehmer ein zufälliges Datum für einen hypothetischen Maßnahmebeginn zugeordnet, welches aus der diskreten Verteilung der Anfangsdaten der Teilnehmer gezogen wird. Falls das derart zugeteilte Datum den institutionellen Voraussetzungen für die Maßnahmenteilnahme zuwiderläuft (die Person also zu diesem Zeitpunkt nicht oder nicht lange genug arbeitslos ist), so wird diese Beobachtung gelöscht. Lechner (2002) weist darauf hin, dass die gelöschten Spells (im Falle seiner Anwendung auf arbeitsmarktpolitische Maßnahmen in der Schweiz 37 % der Nichtteilnehmer) im Durchschnitt kürzer sind als die im Sample. Dies stellt hinsichtlich der Konsistenz kein Problem dar, da im Anschluss das Matching erfolgt und die zeitabhängigen Merkmale dadurch ausbalanciert werden. Das Löschen der Spells dürfte aber implizieren, dass diese Methode gegenüber anderen nicht effizient ist. Bei der vorliegenden Anwendung ist dies allerdings wegen der Größe des Datensatzes von untergeordneter Bedeutung.

Lechner (2002) nutzt außerdem die zeitunabhängigen Kovariaten, um das Anfangsdatum für die Nichtteilnehmer vorherzusagen. Die Anfangsdaten werden dann zufällig aus der auf die zeitunabhängigen Kovariaten bedingten Verteilung der Anfangsdaten der Teilnehmer gezogen. In der Anwendung erhält Lechner bei der Regression der logarithmierten Anfangsdaten auf die zeitunabhängigen Kovariaten ein R^2 von 5 %. Durch die teilweise Vorhersage der Daten müssen nur noch 28 % der Datensätze wegen Unvereinbarkeit mit den Voraussetzungen gelöscht werden. Allerdings ist die Sensitivität der Ergebnisse gegenüber dieser Änderung der Spezifikation gering.

Auf Basis der zugelosten Anfangsdaten können nun für die Nichtteilnehmer alle zeitabhängigen Merkmale und Ergebnisvariablen ermittelt werden. Erstere werden dann zur Schätzung des Propensity Scores verwendet.

c) Aufblähung des Datensatzes

Schließlich schlägt Lechner (1999) eine Methode vor, bei der jedem Nichtteilnehmer systematisch mehrere mögliche Anfangsdaten zugeteilt und die Datensätze entsprechend vervielfältigt werden. Konkret wird für jeden Nichtteilnehmerspell ein Anfangsdatum für jeden Monat innerhalb eines bestimmten Zeitraumes geschaffen und für jedes dieser Anfangsdaten ein eigener Datensatz generiert. Auch hier können wohl nur Anfangsdaten verwendet werden, die mit den institutionellen Voraussetzungen für die Maßnahmenteilnahme vereinbar sind. Bspw. erhält man für einen Nichtteilnehmer, der gut 9 Monate lang arbeitslos und von Anfang an anspruchsberechtigt ist, 9 Datensätze mit Anfangsdatum 1, 2 ... 9 Monaten nach Eintritt in die Arbeitslosigkeit.

Mit dieser Methode erhält Lechner in seiner Anwendung auf die Evaluation von Weiterbildungsmaßnahmen in Ostdeutschland die fragwürdigsten Ergebnisse in dem Sinne, dass durch das Propensity Score Matching der Anteil der Arbeitslosen einen Monat vor der Maßnahme sowie andere Kovariaten schlecht ausbalanciert werden.¹⁴ Dies muss aber nicht auf die Unterlegenheit der Methode hinweisen, sondern kann z. B. daran liegen, dass die Schätzgleichung für den Propensity Score nicht angemessen spezifiziert wurde (vgl. Dehejia, Wahba 2002, Dehejia 2003, Smith, Todd 2003). Dies wäre ggf. anhand von Balancing Tests zu überprüfen. Außerdem ist die verwendete Stichprobe sehr klein (122 Teilnehmer, vgl. Lechner 1999, Tabelle 2).

Für die Zwecke unserer Untersuchung hat diese Methode jedoch einen gravierenden praktischen Nachteil: Durch das Aufblähen des Datensatzes steigt das Datenvolumen um ein Vielfaches an. Da der verwendete Datensatz ohnehin schon sehr umfangreich ist, ist diese Methode somit computertechnisch nicht ohne weiteres zu bewältigen. In dieser Hinsicht ist die im Folgenden beschriebene Methode von Sianesi vorteilhaft, die der eben beschriebenen sehr ähnlich ist.

¹⁴ D. h. für diese Kovariaten gibt es immer noch beachtliche Differenzen zwischen Teilnehmer- und Kontrollgruppe.

d) Stratifizierung nach Dauer der Arbeitslosigkeit

Sianesi (2001) stratifiziert in ihrer Evaluationsstudie zu aktiven Arbeitsmarktprogrammen in Schweden das Sample nach der (diskretisierten) Dauer der Arbeitslosigkeit in Monaten. Im sechsten Stratum sind also bspw. die Teilnehmer, die nach sechs Monaten mit der Maßnahme beginnen sowie alle Arbeitslosen, die nach sechs Monaten Arbeitslosigkeit noch keine Maßnahme begonnen haben. Für diese ist das hypothetische Anfangsdatum dann sechs Monate nach Beginn der Arbeitslosigkeit.

Es werden auch solche Arbeitslose als potenzielle Kontrollpersonen verwendet, die zu einem späteren Zeitpunkt eine Maßnahme beginnen. Der geschätzte ATT besagt dann, welche Wirkung eine Teilnahme zum jetzigen Zeitpunkt gegenüber einem Abwarten hat. Sianesi argumentiert, dass im schwedischen System dieser Effekt interessanter ist, da ein Großteil der Arbeitslosen früher oder später an einer Maßnahme teilnimmt. Für die vorliegende Untersuchung wären ggf. nur diejenigen Arbeitslosen in die Gruppe der potenziellen Kontrollpersonen aufzunehmen, die gar nicht teilnehmen. Gegebenenfalls wäre auch eine Sensitivitätsanalyse mit solchen Kontrollpersonen erforderlich, die an keiner Maßnahme der aktiven Arbeitsmarktpolitik teilnehmen. Auf Basis der im Projekt zur Verfügung stehenden Daten ist dies jedoch nicht möglich.

Für die so gebildeten Strata wird nun jeweils der ATT mittels Propensity Score Matching geschätzt. Diese Methode gibt somit zugleich einen Hinweis darauf, nach welcher Arbeitslosigkeitsdauer die Maßnahme größere Wirkung zeigt. Die Einzeleffekte können anschließend zu einem Durchschnittseffekt aggregiert werden, wobei mit der Anzahl der Teilnehmer im jeweiligen Stratum zu gewichten ist.

Die Methode ist der zuvor beschriebenen Methode der Aufblähung sehr ähnlich. Die Stratifizierung kann nämlich auch derart umgesetzt werden, dass der Datensatz zunächst wie im vorhergehenden Abschnitt aufgebläht und anschließend nach der Dauer der Arbeitslosigkeit gespalten wird. Danach erfolgt das Matching. Ein praktischer Vorteil dieser Vorgehensweise besteht darin, dass das Datenvolumen besser zu handhaben ist.

2.5.2 Balancing Tests

In Abschnitt 2.4.4 wurde darauf hingewiesen, dass beim Propensity Score Matching eine sorgfältige Spezifikation der Schätzgleichung für den Propensity Score erforderlich ist, um die Kovariaten durch das Matching auszubalancieren. Es gibt jedoch keine einheitlichen Kriterien dafür, wann diese Ausbalancierung gelungen ist. Vielmehr werden in der Literatur verschiedenen Balancing Tests angeführt.

Ein häufig angewendeter Test ist die Betrachtung der standardisierten Differenz, die von Rosenbaum und Rubin (1985) eingeführt wurde. Diese ist vor bzw. nach Matching definiert als

$$SDIFF_{vor} = 100 \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{(V_1(X_1) + V_0(X_0))/2}} \quad SDIFF_{nach} = 100 \frac{\bar{X}_{1M} - \bar{X}_{0M}}{\sqrt{(V_1(X_1) + V_0(X_0))/2}} \quad (20)$$

Dabei sind \bar{X}_1 und \bar{X}_0 die Mittelwerte der betrachteten Kovariaten in den gesamten (ungematchten) Teilnehmer- und Nichtteilnehmer-Samples, \bar{X}_{1M} und \bar{X}_{0M} diejenigen in den gematchten Samples. $V_1(X_1)$ und $V_0(X_0)$ sind die Varianzen in den gesamten (ungematchten) Teilnehmer- und Nichtteilnehmer-Samples. Es wird somit in beiden Fällen mit den Stichprobenvarianzen vor Matching normiert, so dass die standardisierten Differenzen vor und nach Matching vergleichbar sind. Problematisch ist an diesem Test, dass es kein formales Kriterium dafür gibt, welche Größe der standardisierten Differenz noch akzeptabel ist. Rosenbaum und Rubin (1985) bezeichnen einen Wert von 20 als groß („large“).

Smith und Todd (2003) schlagen zwei weitere Balancing Tests vor. Zum einen führen sie einen Hotelling – T^2 – Test auf die Null-Hypothese durch, dass die Kovariaten den gleichen Mittelwert bei den Teilnehmern und den gematchten Nichtteilnehmern haben. Sie weisen darauf hin, dass dieser Test strenggenommen berücksichtigen müsste, dass die Matching-Gewichte zufällig sind. In welche Richtung sich die Nichtberücksichtigung auswirkt, lässt sich nicht sagen.

Außerdem schlagen Smith und Todd (2003) einen Test vor, bei dem die folgende Regression für jede Kovariate geschätzt wird:

$$\begin{aligned} X_k = & \beta_0 + \beta_1 \hat{P}(X) + \beta_2 \hat{P}(X)^2 + \beta_3 \hat{P}(X)^3 + \beta_4 \hat{P}(X)^4 \\ & + \beta_5 D + \beta_6 D \hat{P}(X) + \beta_7 D \hat{P}(X)^2 + \beta_8 D \hat{P}(X)^3 + \beta_9 D \hat{P}(X)^4 + \eta \end{aligned} \quad (21)$$

Dann wird die gemeinsame Nullhypothese getestet, dass die Koeffizienten aller Terme, die D (Dummy für die Maßnahmenteilnahme) involvieren, gleich null sind. Problematisch ist bei diesem Test, dass nicht klar ist, bis zu welcher Ordnung die Polynome berücksichtigt werden sollen.

Ein weiterer in der Literatur verwendeter Test ist der two-sample-T-Test auf Abweichung in den Mittelwerten zwischen der Gruppe der Teilnehmer und derjenigen der Nichtteilnehmer.

Insbesondere ist es sinnvoll zu testen, ob sich die Mittelwerte der Ergebnisvariablen in der Treatmentgruppe und der gematchten Kontrollgruppe vor dem Treatment signifikant unterscheiden. Ist dies der Fall, so ist die Selektionsverzerrung offenbar noch nicht beseitigt, sofern der Treatmentzeitpunkt so definiert ist, dass vor diesem Zeitpunkt noch keine Wirkungen der Maßnahme auftreten können. Umgekehrt kann man allerdings nicht sagen, dass die Selektionsverzerrung behoben ist, wenn es keine signifikanten Differenzen mehr gibt. Es ist nämlich möglich, dass die Selektionsverzerrung auf unbeobachtbaren Merkmalen beruht, die im Zeitablauf nicht konstant sind oder zumindest keinen konstanten Einfluss auf die Beschäftigungswahrscheinlichkeit aufweisen.

3 Datenbasis

Die mikroökonomische Analyse wird im vorliegenden Bericht für die Monate Mai und Juni 2003 durchgeführt. Als Datenbasis kann somit weitgehend der Datensatz verwendet werden, der für das Projektmonitoring dieser Monate erstellt wurde (vgl. Teil I des Forschungsberichtes). Es ist noch zu überlegen, wie bei der Auswahl des analysierten Samples mit der Anspruchsberechtigung umzugehen ist. Eine Möglichkeit ist es, diejenigen Spells zu löschen, bei denen die Person zum Referenzzeitpunkt nicht anspruchsberechtigt ist, da diese ohnehin keinen VGS erhalten können und somit als Kontrollpersonen für die VGS-Bezieher nicht in Frage kommen.

Allerdings erhalten einige tausend Personen, für die sich aus den Daten keine Anspruchsberechtigung zum Referenzzeitpunkt ergibt, dennoch einen VGS. Zum Teil waren diese Personen nur für wenige Tage innerhalb der letzten vier Monate arbeitslos, zum Teil hatten sie keinen Anspruch auf Leistungen. Dies kann daran liegen, dass die Berechnungen zur Ermittlung der Anspruchsberechtigung ungenau sind, dass es Fehler in den Daten gibt oder dass tatsächlich VGS an nicht Anspruchsberechtigte ausgegeben werden. Diese Personen müsste man konsequenterweise ebenfalls aus der Analyse ausschließen, sodass man keine Aussage über die Wirkung ihrer VGS treffen könnte. Um dies zu vermeiden, kann man stattdessen die nicht Anspruchsberechtigten im Sample belassen. In diesem Fall muss man jedoch das Merkmal „Anspruchsberechtigung“ als Kovariate beim Propensity-Score-Matching verwenden, um insoweit eine Selektionsverzerrung zu vermeiden. Im Folgenden werden beide Möglichkeiten durchgeführt, um die Sensitivität hinsichtlich dieser Sampleauswahl zu überprüfen.

Eine weitere Einschränkung des Samples muss wegen fehlender Daten erfolgen. In Ostdeutschland gibt es 820.334 Spells für Anspruchsberechtigte, von denen 30.996 einen Gutschein erhalten. Entfernt man die Spells mit fehlenden Informationen für Kovariaten, so verbleiben noch 787.386 Spells, wovon bei 29.785 eine Ausgabe erfolgt. In Westdeutschland gibt es 1.538.520 Spells mit Anspruchsberechtigung, davon 35.040 mit Ausgabe. Nach der Entfernung der Missings verbleiben 1.440.354 Spells und 32.600 Ausgaben.

Die Missings beziehen sich fast ausschließlich auf die Variable „Dauer des letzten Beschäftigungsverhältnisses“, die auf den BewA-Daten basiert. Da die Erwerbshistorie ohnehin noch zusätzlich aus der IEB herausgezogen wird, wird im Rahmen der Sensitivitätsanalyse getestet, ob sich die Ergebnisse ändern, wenn man diese Kovariate weglässt und dafür die zusätzlichen Beobachtungen verwendet.

Mehrfachausgaben

Viele der Teilnehmer erhalten nicht zum ersten Mal einen Vermittlungsgutschein. In Tabelle 0 wird dargestellt, wie viele Gutscheine die Teilnehmer in den Monaten Mai und Juni 2003 zuvor schon erhalten hatten. 78 Prozent von ihnen hatten zuvor noch keinen VGS erhalten. In der ökonomischen Analyse werden nur die jeweils ersten VGS innerhalb des

Zweimonatsintervalls betrachtet. Falls ein Arbeitsloser in den Monaten Mai und Juni also zwei VGS erhält, wird der zweite nicht betrachtet. Dies betrifft weniger als 200 Arbeitslose.¹⁵

Tabelle 0: Mehrfachausgaben von Vermittlungsgutscheinen: Anzahl von zuvor erhaltenen VGS für die Teilnehmer im Mai und Juni 2003

Zuvor erhaltene VGS	Anzahl	Prozent
keiner	51.548	78,06
1	6.894	10,44
2	3.573	5,41
3	1.862	2,82
4	1.063	1,61
5 und mehr	1.096	1,66
Summe	66.036	100,00

4 Ergebnisse

In diesem Teil werden die Ergebnisse der mikroökonomischen Analyse dargestellt, bei der die Individualeffekte der Vermittlungsgutscheine abgeschätzt werden. Alle Schätzungen wurden in dem Ökonometrieprogramm Stata umgesetzt. Die zentrale Zielgröße ist der Verbleib in regulärer Beschäftigung, wie bereits im Rahmen des Projektmonitoring diskutiert (vgl. Teil I des Forschungsberichtes). Es erfolgt außerdem eine Disaggregation dieser Größe in die Determinanten Übergänge in Beschäftigung und Dauer der Beschäftigungsverhältnisse. Alle Untersuchungen beziehen sich auf die VGS, die in den Monaten Mai und Juni 2003 an Arbeitslose ausgegeben wurden. Aus Gründen der Datenbereitstellung können im Augenblick die Effekte von VGS, deren Ausgabe später erfolgte, nicht evaluiert werden. Für die Evaluation erscheint es nämlich erforderlich, für die 6 Monate nach der Ausgabe des Gutscheins Informationen über den Beschäftigtenstatus des jeweiligen Arbeitslosen zu haben. Wie in Teil I des Forschungsbericht-

¹⁵ Bei der Analyse der Eingliederungswirkungen wurde u. a. danach differenziert, ob der Teilnehmer vor dem betrachteten VGS schon einmal einen erhalten hatte. Es zeigt sich, dass die „Erstteilnehmer“ einen überdurchschnittlich hohen Treatment Effekt aufweisen, während sich für die übrigen Teilnehmer kein signifikant positiver Effekt ergibt. Dieses Ergebnis ist offensichtlich auf einen Selektionseffekt zurückzuführen: Arbeitslose, die den VGS zum wiederholten Male erhalten, sind entweder zuvor nicht vermittelt worden oder schon wieder arbeitslos. Es handelt sich also um Personen, denen der VGS zuvor relativ wenig gebracht hat.

tes dargelegt, sind die Beschäftigtendaten im Augenblick nur bis Dezember 2003 mit ausreichender Genauigkeit verfügbar.

Die Analyse wird außerdem aufgrund der unterschiedlichen Arbeitsmarktsituationen für Ost- und Westdeutschland getrennt durchgeführt. Neben den durchschnittlichen Treatment Effekten für die gesamten Samples werden auch Treatment Effekte für Subsamples geschätzt, um die Heterogenität der Effekte abzubilden. Eine Differenzierung erfolgt nach den Merkmalen Art des VGS, Geschlecht, regionaler Vergleichstyp und Propensity Score.

Zur Berücksichtigung der zeitvariierenden Variablen wird das Verfahren von Lechner gewählt, bei dem hypothetische Ausgabedaten für die Nichtteilnehmer zufällig aus der Verteilung der Ausgabedaten für die VGS-Bezieher gezogen werden (vgl. Abschnitt 2.5.1, Punkt b sowie Teil I des Forschungsberichtes).

Die ersten Schätzungen erfolgen für das Sample der (in den Daten identifizierten) Anspruchsberechtigten. Bei der Darstellung der Ergebnisse wird wie folgt vorgegangen. Zunächst werden in Abschnitt 4.1 die Ergebnisse der Probit-Schätzung des Propensity Scores präsentiert. Anschließend wird für die verschiedenen Matching-Varianten anhand von Balancing Tests dargestellt, inwieweit die Kovariaten und die Outcome-Variable vor dem Treatment ausbalanciert werden (Abschnitt 4.2). In Abschnitt 4.3 werden die auf Basis des Matchings geschätzten Treatment Effekte für das gesamte Sample dargestellt, in Abschnitt 4.4 diejenigen für bestimmte Subsamples. Anschließend werden noch die Ergebnisse der Sensitivitätsanalyse hinsichtlich der Sampleauswahl (Abschnitt 4.5) aufgezeigt.

4.1 Schätzung des Propensity Scores

In der Probitschätzung wird zunächst bestimmt, welche Charakteristika einen Einfluss auf die Wahrscheinlichkeit der Maßnahmenteilnahme haben und wie groß dieser Einfluss im Einzelnen ist. In der Analyse wurden allgemeine sozioökonomische Merkmale wie Alter, Familienstand, Kinder, Behinderungen und andere gesundheitliche Einschränkungen berücksichtigt. Die Qualifikation der Arbeitslosen wird zum einen über die formalen Abschlüsse der Schul- und Berufsausbildung und zum anderen über eine Einschätzung der Arbeitsamtsmitarbeiter und über das Vorliegen von Be-

rufserfahrung erfasst. Da gleichzeitig für den Arbeitsagenturbezirk und damit für die regionale Arbeitsmarktsituation kontrolliert wird, ist damit implizit auch das individuelle Verhältnis von vorhandener zu nachgefragter Qualifikation erfasst.

Des Weiteren gehen in die Schätzung Vermittlungswünsche der Arbeitslosen bezüglich des Berufsfeldes und der Arbeitszeit ein, welche die Flexibilität der Arbeitslosen widerspiegeln sollen. Wichtig für die Bestimmung der Determinanten der Maßnahmenteilnahme sollten auch Variablen sein, die den letzten Kontakt des Arbeitslosen mit dem Arbeitsmarkt beschreiben. Dazu gehören Merkmale wie die Dauer und das Entgelt der letzten Beschäftigung und ob diese durch den Arbeitgeber beendet wurde oder ob ein befristeter Vertrag auslief. Außerdem werden die Teilnahme an ABM, SAM und REHA, sowie die Zahlung von Eingliederungszuschüssen erfasst.

Zudem sollten die Motivation und die Verfügbarkeit des Arbeitslosen die Teilnahmewahrscheinlichkeit beeinflussen. Dazu verwenden wir Informationen, ob und wie oft in den letzten 5 Jahren Abmeldungen wegen fehlender Verfügbarkeit, wegen Krankheiten und wegen verhängter Säumniszeiten oder Sperrzeiten vorliegen. Diese Informationen dürften auch die Beschäftigungswilligkeit widerspiegeln.

Weiter handelt es sich um Variablen der Erwerbsbiographie, die Indikatoren für die Anbindung und die Anstrengungen auf dem Arbeitsmarkt in den letzten 5 Jahren darstellen. In dieser Schätzung wird zunächst die Anzahl der Tage in Beschäftigung, Arbeitslosigkeit, FbW und ABM aus den letzten 5 Jahren in verschiedenen Zeiträumen erfasst (beispielsweise die Anzahl der Tage im letzten Halbjahr vor dem Analysedatum)¹⁶.

Die bisherige Dauer der aktuellen Arbeitslosigkeit geht (ebenso wie das Alter) über eine Splinefunktion in einer sehr flexiblen Form in die Bestimmung der Teilnahmewahrscheinlichkeit ein. Bei einer Splinefunktion kann der Einfluss der berücksichtigten Merkmale nicht einzeln abgelesen werden, sondern muss durch Addition gebildet werden. Wenn man bspw. den marginalen Einfluss einer Verlängerung der Arbeitslosigkeitsdauer um ei-

¹⁶ Darüber hinaus soll auch die Teilnahme an Maßnahmen der Arbeitsämter in der nächsten Schätzung berücksichtigt werden.

nen Tag bestimmen will, und die Arbeitslosigkeitsdauer zwischen 4 und 5 Monaten liegt, so muss die Summe aus den geschätzten Koeffizienten für die Teilfunktionen zwei, drei und vier gebildet werden. Genauso verhält es sich beim Alter. Wenn man den richtigen Koeffizienten für einen 31-Jährigen bestimmen will, müssen die Koeffizienten der Variablen Alter 18, Alter 25 und Alter 30 addiert werden.

Der Vorteil einer solchen Splinefunktion gegenüber der Berücksichtigung von einer stetigen Variablen liegt in der Flexibilität. Man unterstellt bei einer solchen Funktion keinen stetigen linearen oder quadratischen Zusammenhang zwischen der Teilnahmewahrscheinlichkeit und dem Merkmal. Bei einem linearen Zusammenhang kann man eine Gerade zwischen der Ausprägung des Merkmals und der Teilnahmewahrscheinlichkeit ziehen. Diese hat die Steigung des geschätzten Koeffizienten aus einer Schätzung, in der das Merkmal in stetiger Form eingeht. Bei der Splinefunktion wird das Merkmal in Intervalle geteilt, zum Beispiel kann die Länge eines Intervalls einem Monat Arbeitslosigkeit entsprechen. Es wird in diesen Intervallen jeweils der Anstieg einer Geraden bestimmt, die den Zusammenhang zwischen Teilnahmewahrscheinlichkeit und Merkmal in diesem Intervall wiedergibt. Dadurch kann der Einfluss des Merkmals viel differenzierter erfasst werden. Schließlich gehen in die Analyse auch noch regionale Merkmale ein, die den Arbeitsmarkt vor Ort beschreiben sollen. An dieser Stelle werden nicht alle Merkmale im Detail besprochen.¹⁷ Andere Variablen werden der Vollständigkeit halber aufgeführt, haben aber keinen signifikanten Einfluss¹⁸ auf die Teilnahmewahrscheinlichkeit.

¹⁷ Die Koeffizienten der Berufsfelder und der Arbeitsämterdummys sowie Interaktionsterme werden in der Tabelle mit der Probitschätzung aus Platzgründen nicht aufgeführt, sind aber auf Nachfrage erhältlich.

¹⁸ Ein Koeffizient ist auf einem Niveau von 5 % signifikant, wenn z kleiner als -2 bzw. größer als 2 ist.

Tabelle 1: Probitschätzung des Propensity Scores

	Ost		West	
Pseudo R-squared			0.0736	0.1107
Percentage correctly predicted Teilnehmer			69,52 %	70,97 %
Percentage correctly predicted Nichtteilnehmer			62,98 %	69,20 %
Variable	Koeffizient (marginaler Effekt)	z	Koeffizient (marginaler Effekt)	z
<i>Allgemeine sozioökonomische Merkmale</i>				
Splinefunktion				
Alter 0	-0,0010647	-3,72	0,000781	5,52
Alter 25	-0,0003872	-1,63	-0,0000929	-0,84
Alter 30	0,0002942	1,35	0,000085	0,85
Alter 35	0,0000652	0,35	0,0001222	1,35
Alter 40	-0,0000126	-0,07	-0,0002055	-2,22
Alter 45	-0,0009666	-5,44	-0,0002842	-2,85
Alter 50	-0,0026141	-11,89	-0,0012937	-11,03
Alter 55	-0,0072639	-15,69	-0,0037111	-15,01
Alter 60	-0,0041719	-1,85	-0,0036331	-3,73
Ausländer	-0,007183	-4,54	-0,0012022	-4,3
Frau	0,0006294	0,83	-0,0005544	-1,77
Allein lebend	Referenz			
In Partnerschaft nicht verheiratet	0,0036635	4,33	0,0015601	2,86
allein erziehend	-0,0045811	-2,29	0,0013644	1,27
verheiratet	0,0066559	11,45	0,0014238	5,07
Frau * (in Partnerschaft nicht verheiratet)	-0,0073403	-6,14	-0,0023423	-2,95
Frau * (allein erziehend)	0,0037979	1,59	-0,0012558	-1,15
Frau * (verheiratet)	-0,0096193	-12,07	-0,0040927	-10,24
Kinder 0-3 Jahre	-0,0216069	-1,27	0,0123629	0,63
Kinder 3-7 Jahre	-0,0018466	-2,41	-0,0006094	-1,61
Kinder 7-14 Jahre	-0,0007433	-1,23	-0,000426	-1,27
Gesundheitszustand:				
Keine gesundheitlichen Einschränkungen	Referenz			
Behinderungsgrad (BG) 80 %	-0,0148563	-5,05	-0,006571	-7,01
BG 50 %	-0,0101988	-8,02	-0,0049864	-10,03
BG 30 %, nicht gleichgestellt	-0,0027611	-1,81	-0,0014772	-2,59
BG 30 %, gleichgestellt	0,0006178	0,34	-0,0015947	-1,5
Sonst. gesundheitliche Einschränkungen	-0,0032585	-6,03	-0,0023596	-9,15
<i>Qualifikation</i>				
Schulabschluss:				
Ohne Abschluss	Referenz			
Hauptschulabschluss	0,0016704	2,07	0,0003371	1,07
Mittlere Reife	0,0017754	2,19	0,0017687	4,61
Fachhochschulreife	-0,0013129	-0,79	0,0014258	2,57
Abitur	0,0002701	0,21	0,0004821	0,93
Berufliche Ausbildung:				
Keine Ausbildung	Referenz			
Betriebliche Ausbildung	0,0088969	7,11	0,0035428	5,74
Außerbetriebliche Ausbildung	0,0131557	6,69	0,0105965	7,71
Berufsfachschule, Fachschule	0,0155769	4,43	0,0051233	3,62

Universität	-0,0029178	-0,63	0,0038558	2,41
Fachhochschule	0,0169173	2,7	0,0084369	3,77
Qualifikation nach Einschätzung des Agenturmitarbeiters:				
Kräfte mit und ohne Fachkenntnisse	Referenz			
Spitzenkräfte	0,0033099	0,28	-0,0010207	-0,26
Kräfte mit Hochschulniveau	-0,0001391	-0,07	0,0006674	0,75
Kräfte mit Fachhochschulniveau	0,0065662	3,42	0,0021943	2,4
Kräfte mit Fachschulniveau	0,0083749	6	0,0049377	7,19
Fachkräfte	0,0055505	12,67	0,0022955	8,84
Sonst. Kräfte	0,0087526	5,09	0,0035501	4,42
Berufserfahrung im gewünschten Beruf (Einschätzung AA-Mitarbeiter)				
	0,0043374	3,42	0,0009569	1,92
Jobwechsel	0,0010719	2,83	-0,0002515	-1,35
<i>Zukünftige Beschäftigung: Arbeitszeitwunsch</i>				
Nur Vollzeit	Referenz			
Nur Teilzeit	-0,0069045	-5,03	-0,0039116	-9,54
Zukünftige Arbeitszeit egal	0,0006357	0,72	0,0000315	0,06
Vollzeit, auch Schicht	0,0039786	4,11	0,0056397	5,63
Sonst. Arbeitszeit	-0,0028782	-0,14	0,0006041	0,15
<i>Letzter Kontakt mit AM</i>				
Letzte Beschäftigungsdauer (M)	-0,0000849	-9,87	-0,0000292	-7,23
Letzte Beschäftigungsdauer (M), quadriert	1,24E-07	4,77	2,11E-08	1,71
Dummy positives Tagesentgelt	-0,005597	-7,14	0,0006599	2,46
Tagesentgelt bei letzter Beschäftigung (€)	0,0002005	7,51	6,02E-06	1,71
Tagesentgelt bei letzter Beschäftigung (€), Quadrat	-1,16E-06	-4,69	-5,97E-10	-0,68
Teilnahme an Maßnahmen der AAMP vor Arbeitslosigkeit:				
ABM	-0,005072	-9,12	-0,0020245	-4,15
SAM	-0,0033037	-4,44	-0,0020994	-1,92
Eingliederungszuschüsse	-0,0006952	-0,56	0,0013247	1,54
REHA	-0,0011446	-1,25	0,0011235	2,21
Keine der genannten	Referenz			
Grund für Beendigung des letzten Arbeitsverhältnisses:				
Kündigung durch Arbeitgeber	0,0018347	3,9	0,0013615	6,16
Befristeter Vertrag	-0,0019765	-3,94	0,0017171	5,58
Sonst. Grund	Referenz			
<i>Abmeldungen aus Arbeitslosigkeit</i>				
Krankmeldungen, Dummy	0,0028689	5,97	0,0006507	2,4
Krankmeldungen, Anzahl	0,0006086	4,31	0,0001196	1,75
Fehlende Verfügbarkeit, Dummy	-0,0011746	-1,12	0,000405	1,07
Fehlende Verfügbarkeit, Anzahl	-0,0021378	-3,17	-0,0009322	-4,55
Sperrzeiten, Dummy	-0,0033461	-0,67	-0,0000827	-0,05
Sperrzeiten, Anzahl	0,0042646	0,88	-0,0009508	-0,63
Säumniszeiten, Dummy	0,0000481	0,01	-0,0011977	-0,88
Säumniszeiten, Anzahl	-0,0023017	-0,79	-0,0006538	-0,54
<i>Leistungsbezug</i>				
Leistungsbezug, Höhe	0,000191	2,18	0,0002846	7,69
Leistungsbezug, Höhe, Quadrat	-1,74E-07	-0,09	-4,59E-06	-7,23

Erwerbshistorie

Splinefunktion für Dauer des aktuellen
Arbeitslosigkeitsspiels:

Arbeitslosigkeitsdauer 2 Monate	0,0000904	0,52	-0,0002621	-3,04
Arbeitslosigkeitsdauer 3 Monate	-0,0006729	-4,7	-0,0002282	-3,19
Arbeitslosigkeitsdauer 4 Monate	0,0015524	14,94	0,0006812	13,6
Arbeitslosigkeitsdauer 5 Monate	-0,0007881	-18,12	-0,0002801	-13,5
Arbeitslosigkeitsdauer 6 Monate	0,0000884	1,69	0,0000468	1,92
Arbeitslosigkeitsdauer 7 Monate	0,000221	5,18	0,0000492	2,32
Arbeitslosigkeitsdauer 8 Monate	-0,0001097	-2,29	-0,0000391	-1,66
Arbeitslosigkeitsdauer 9 Monate	-0,0001066	-1,85	-0,0000203	-0,73
Arbeitslosigkeitsdauer 10 Monate	0,0000541	0,98	-6,93E-07	-0,03
Arbeitslosigkeitsdauer 11 Monate	0,0000324	0,54	0,0000231	0,79
Arbeitslosigkeitsdauer 12 Monate	-4,66E-06	-0,07	0,0000132	0,41
Arbeitslosigkeitsdauer 13 Monate	0,0000231	0,35	0,0000176	0,53
Arbeitslosigkeitsdauer 14 Monate	0,0000295	0,42	-0,0000746	-2
Arbeitslosigkeitsdauer 15 Monate	-0,00014	-1,95	0,0000303	0,78
Arbeitslosigkeitsdauer 16 Monate	6,08E-06	0,08	-0,0000481	-1,16
Arbeitslosigkeitsdauer 17 Monate	-0,0000806	-1,07	0,0000206	0,5
Arbeitslosigkeitsdauer 18 Monate	0,0001733	2,24	0,0000296	0,69
Arbeitslosigkeitsdauer 19 Monate	-0,0001708	-2,14	-0,000014	-0,32
Arbeitslosigkeitsdauer 20 Monate	0,0001017	1,14	0,0000237	0,5
Arbeitslosigkeitsdauer 21 Monate	-0,0001498	-1,56	-0,0000514	-1,02
Arbeitslosigkeitsdauer 22 Monate	0,0001222	1,24	-0,000015	-0,28
Arbeitslosigkeitsdauer 23 Monate	-0,0000423	-0,4	-0,0000315	-0,54
Arbeitslosigkeitsdauer 24 Monate	0,0000773	0,75	0,0001134	1,94
Arbeitslosigkeitsdauer 36 Monate	-0,0000563	-0,73	-0,0001418	-3,2

Informationen über die Zeit vor dem
Referenzzeitpunkt (RZ)

Bis 25 % der Tage in Arbeitslosigkeit in den 6 Monaten (M) vor RZ	0,0026595	2,92	0,0005954	1,45
25-50 % der Tage in Arbeitslosigkeit in den 6 M vor RZ	0,0047416	3,54	0,0019605	3,39
50-75 % der Tage in Arbeitslosigkeit in den 6 M vor RZ	0,0028761	0,79	-0,0003601	-0,26
75-100 % der Tage in Arbeitslosigkeit in den 6 M vor RZ	-0,0038601	-0,57	-0,0007758	-0,33
Bis 25 % der Tage in Arbeitslosigkeit in 12-6 M vor RZ	0,0058992	6,17	0,0026529	5,91
25-50 % der Tage in Arbeitslosigkeit 12-6 M vor RZ	0,0079145	6,8	0,0035389	6,67
50-75 % der Tage in Arbeitslosigkeit 12-6 M vor RZ	0,0065831	5,05	0,0023966	4,19
75-100 % der Tage in Arbeitslosigkeit 12-6 M vor RZ	0,003929	3,12	0,002057	3,66
Bis 25 % der Tage in Arbeitslosigkeit 2-1 Jahre vor RZ	0,0043473	5,82	0,0016532	4,17
25-50 % der Tage in Arbeitslosigkeit 2-1 Jahre vor RZ	0,0016594	1,89	0,0003233	0,75
50-75 % der Tage in Arbeitslosigkeit 2-1 Jahre vor RZ	0,0017392	1,75	0,0007385	1,54
75-100 % der Tage in Arbeitslosigkeit 2-1 Jahre vor RZ	-0,0001096	-0,1	0,0008656	1,66

Bis 25 % der Tage in Arbeitslosigkeit 3-2 Jahre vor RZ	0,0033492	4,77	0,0015669	4,04
25-50 % der Tage in Arbeitslosigkeit 3-2 Jahre vor RZ	0,0025643	3,06	0,0011783	2,81
50-75 % der Tage in Arbeitslosigkeit 3-2 Jahre vor RZ	0,0027255	2,84	0,0004218	0,92
75-100 % der Tage in Arbeitslosigkeit 3-2 Jahre vor RZ	0,0026637	2,66	0,0013016	2,8
Bis 25 % der Tage in Arbeitslosigkeit 4-3 Jahre vor RZ	0,002036	2,78	0,0006853	1,8
25-50 % der Tage in Arbeitslosigkeit 4-3 Jahre vor RZ	0,000964	1,21	-0,0000311	-0,08
50-75 % der Tage in Arbeitslosigkeit 4-3 Jahre vor RZ	0,0001204	0,15	3,52E-07	0
75-100 % der Tage in Arbeitslosigkeit 4-3 Jahre vor RZ	0,0018877	2,24	0,0006463	1,53
Bis 25 % der Tage in Arbeitslosigkeit 5-4 Jahre vor RZ	0,0014001	1,94	0,0005568	1,55
25-50 % der Tage in Arbeitslosigkeit 5-4 Jahre vor RZ	0,0013874	1,7	-0,0002954	-0,74
50-75 % der Tage in Arbeitslosigkeit 5-4 Jahre vor RZ	-0,0007698	-1,02	-0,0001665	-0,41
75-100 % der Tage in Arbeitslosigkeit 5-4 Jahre vor RZ	0,001642	2,37	0,0013414	3,92
Bis 25 % der Tage in Beschäftigung in den 6 M vor RZ	-0,0008339	-0,32	-0,0003569	-0,22
25-50 % der Tage in Beschäftigung in den 6 M vor RZ	-0,0009415	-0,24	-0,001171	-0,59
50-75 % der Tage in Beschäftigung in den 6 M vor RZ	0,0018077	0,47	0,0007571	0,38
75-100 % der Tage in Beschäftigung in den 6 M vor RZ	0,0005046	0,13	0,0003345	0,17
Bis 25 % der Tage in Beschäftigung in den 12-6 M vor RZ	0,0032379	3,73	0,0014276	3,32
25-50 % der Tage in Beschäftigung in den 12-6 M vor RZ	0,0014302	1,21	0,0007171	1,28
50-75 % der Tage in Beschäftigung in den 12-6 M vor RZ	0,0028785	2,28	0,00101	1,69
75-100 % der Tage in Beschäftigung in den 12-6 M vor RZ	0,0013052	1,07	0,000555	0,95
Bis 25 % der Tage in Beschäftigung 2-1 Jahre vor RZ	0,0033896	4,9	0,0014101	4,26
25-50 % der Tage in Beschäftigung 2-1 Jahre vor RZ	0,0020811	2,29	0,0012947	2,94
50-75 % der Tage in Beschäftigung 2-1 Jahre vor RZ	0,0003515	0,35	0,0005198	1,04
75-100 % der Tage in Beschäftigung 2-1 Jahre vor RZ	-0,0044408	-4,04	-0,0014216	-2,62
Bis 25 % der Tage in Beschäftigung 3-2 Jahre vor RZ	0,0017416	2,74	0,0013711	4,43
25-50 % der Tage in Beschäftigung 3-2 Jahre vor RZ	-0,0003418	-0,42	0,0004285	1,07
50-75 % der Tage in Beschäftigung 3-2 Jahre vor RZ	-0,0017561	-2	-0,000271	-0,59
75-100 % der Tage in Beschäftigung 3-2 Jahre vor RZ	-0,0061542	-6,51	-0,0022461	-4,54

Bis 25 % der Tage in Beschäftigung 4-3 Jahre vor RZ	0,0017663	2,84	0,0012259	3,8
25-50 % der Tage in Beschäftigung 4-3 Jahre vor RZ	0,0011913	1,5	0,0011145	2,68
50-75 % der Tage in Beschäftigung 4-3 Jahre vor RZ	-0,0016491	-1,89	0,0002825	0,6
75-100 % der Tage in Beschäftigung 4-3 Jahre vor RZ	-0,0043679	-5,16	-0,0014694	-3,19
Bis 25 % der Tage in Beschäftigung 5-4 Jahre vor RZ	0,0012545	2,17	0,0012857	3,91
25-50 % der Tage in Beschäftigung 5-4 Jahre vor RZ	0,0022187	2,79	0,0016426	3,98
50-75 % der Tage in Beschäftigung 5-4 Jahre vor RZ	-0,000925	-1,06	0,0020423	4,38
75-100 % der Tage in Beschäftigung 5-4 Jahre vor RZ	-0,002605	-3,37	-0,000203	-0,5
Bis 25 % der Tage in ABM in den 6 M vor RZ	0,0066864	4,28	0,0013003	0,96
25-50 % der Tage in ABM in den 6 M vor RZ	0,0042251	2,17	0,0028791	1,9
50-75 % der Tage in ABM in den 6 M vor RZ	0,010222	2,83	0,0064255	2,66
75-100 % der Tage in ABM in den 6 M vor RZ	0,0118461	2,64	0,002636	1,14
Bis 25 % der Tage in ABM in den 12-6 M vor RZ	0,0021536	1,28	0,0013338	0,99
25-50 % der Tage in ABM in den 12-6 M vor RZ	0,0035801	2,24	0,0020356	1,57
50-75 % der Tage in ABM in den 12-6 M vor RZ	0,0019901	1,19	0,0022114	1,71
75-100 % der Tage in ABM in den 12-6 M vor RZ	0,0009173	0,55	0,0026622	2,08
Bis 25 % der Tage in ABM 2-1 Jahre vor RZ	0,0011051	0,97	-0,0002025	-0,24
25-50 % der Tage in ABM 2-1 Jahre vor RZ	-0,0005284	-0,48	0,000016	0,02
50-75 % der Tage in ABM 2-1 Jahre vor RZ	-0,0006507	-0,5	-0,0013337	-1,42
75-100 % der Tage in ABM 2-1 Jahre vor RZ	0,0004777	0,32	0,0007168	0,67
Bis 25 % der Tage in ABM 3-2 Jahre vor RZ	0,0008932	0,75	0,0022902	2,44
25-50 % der Tage in ABM 3-2 Jahre vor RZ	-0,0007719	-0,7	0,0012678	1,55
50-75 % der Tage in ABM 3-2 Jahre vor RZ	0,000022	0,02	0,0023823	2,6
75-100 % der Tage in ABM 3-2 Jahre vor RZ	-0,000523	-0,43	0,0015089	1,59
Bis 25 % der Tage in FbW in den 6 M vor RZ	-0,0003206	-0,4	0,0010951	2,72
25-50 % der Tage in FbW in den 6 M vor RZ	0,0059983	3,7	0,0039988	5,35
50-75 % der Tage in FbW in den 6 M vor RZ	0,0041873	0,9	-0,0013907	-0,62
75-100 % der Tage in FbW in den 6 M vor RZ	-0,0081145	-1,62	-0,004395	-1,91

Bis 25 % der Tage in FbW in den 12-6 M vor RZ	0,0011933	1,47	0,0021237	4,6
25-50 % der Tage in FbW in den 12-6 M vor RZ	0,0058265	3,89	0,0025473	3,41
50-75 % der Tage in FbW in den 12-6 M vor RZ	0,0065557	3,11	0,0062891	5,48
75-100 % der Tage in FbW in den 12-6 M vor RZ	0,011008	6,05	0,0072018	7,34
Bis 25 % der Tage in FbW 2-1 Jahre vor RZ	0,0018823	2,77	0,0022514	5,6
25-50 % der Tage in FbW 2-1 Jahre vor RZ	0,0051988	4,16	0,0028856	4,1
50-75 % der Tage in FbW 2-1 Jahre vor RZ	0,004628	3,18	0,0035929	3,92
75-100 % der Tage in FbW 2-1 Jahre vor RZ	0,0044848	2,8	0,0028222	2,89
Bis 25 % der Tage in FbW 3-2 Jahre vor RZ	0,0015566	2,05	0,0017269	3,86
25-50 % der Tage in FbW 3-2 Jahre vor RZ	0,0039156	3,1	0,0020083	2,93
50-75 % der Tage in FbW 3-2 Jahre vor RZ	0,002565	1,79	0,0034707	3,89
75-100 % der Tage in FbW 3-2 Jahre vor RZ	0,0062604	3,89	0,0034851	3,35
<i>Regionale Indikatoren (auf Ebene des Arbeitsagenturbezirks, Quoten)</i>				
Arbeitslose/Erwerbstätige	0,001814	3,12	2,87E-06	6,07
Gemeldete Stellen/Erwerbstätige	0,0157701	1,95	-0,0024523	-4,07
Kurzarbeiter/Erwerbstätige			0,0275061	15,59
ABM/Erwerbstätige	-0,0238592	-6,53		
FbW/Erwerbstätige	-0,0062489	-2,8	-0,0520825	-13,71
Arbeitslose/Mitarbeiter	-0,000087	-2,01	-0,0000648	-7,97
<i>Berücksichtigung saisonaler Effekte</i>				
Ausgabe im Juni	0,0028653	8,3	-0,0004701	-2,65

In der Tabelle 1 sind die Ergebnisse der Probitschätzung für Ost- und Westdeutschland aufgeführt. Bei den allgemeinen Merkmalen und den Variablen, die die Qualifikation der Arbeitslosen beschreiben, ist der geschätzte Einfluss auf die Teilnahmewahrscheinlichkeit in Ost- und Westdeutschland sehr ähnlich. So weisen Behinderte und Ausländer eine geringe Wahrscheinlichkeit auf, sich einen Vermittlungsgutschein ausstellen zu lassen.

Frauen haben im Vergleich zu Männern keine höhere Wahrscheinlichkeit, wobei dieser Einfluss zum Teil durch die Interaktionsterme mit dem Familienstand aufgefangen wird. Bezüglich des Familienstandes zeigt die Schätzung, dass verheiratete und nichtverheiratete, aber in Partnerschaft lebende Arbeitslose eher auf die Vermittlungstätigkeit Dritter zurückgrei-

fen als Singles. Kinder wirken sich in Ost- und Westdeutschland anders auf die Teilnahmewahrscheinlichkeit aus. Während es in Westdeutschland keine Unterschiede zwischen Arbeitslosen mit und ohne Kinder (unabhängig von ihrem Alter) bezüglich der Ausstellung von Vermittlungsgutscheinen gibt, ist die Teilnahmewahrscheinlichkeit in Ostdeutschland bei Eltern von Kindern im Vorschulalter geringer als bei kinderlosen Arbeitslosen. Möglicherweise schätzen ostdeutsche Eltern ihre Beschäftigungschancen pessimistischer ein und nehmen deswegen die Möglichkeit der Vermittlung durch Dritte nicht wahr. Mangelnde Betreuungsmöglichkeiten in Kindertagesstätten dürften vermutlich nicht der Grund sein, da dies eher in Westdeutschland auftritt. Behinderte und sonstig gesundheitlich eingeschränkte Arbeitslose haben insgesamt eine geringere Wahrscheinlichkeit der Maßnahmenteilnahme. Da auch sie ihre Beschäftigungsaussichten eher schlechter einschätzen dürften, ist dieses Ergebnis auch zu erwarten gewesen.

Bei den Merkmalen, die die Qualifikation der Arbeitslosen beschreiben, zeichnet sich in der Schätzung bezüglich der Teilnahmewahrscheinlichkeit kein eindeutiger Trend ab. Während sich ostdeutsche Arbeitslose mit einem Hauptschulabschluss oder der Mittleren Reife bzw. westdeutsche mit der Mittleren oder der Fachhochschulreife eher einen Vermittlungsgutschein ausstellen lassen als jene ohne Schulabschluss, gibt es zwischen denen ohne Abschluss und Arbeitslosen mit Abitur keinen signifikanten Unterschied. Was die berufliche Ausbildung angeht, greifen alle Arbeitslosen mit einem Abschluss, egal ob es sich dabei um einen Abschluss einer betrieblichen Ausbildung oder einer Hochschule handelt, eher auf Vermittlungsgutscheine zurück als jene ohne Abschluss. Dabei kann man nicht feststellen, dass die Wahrscheinlichkeit mit dem Bildungsniveau steigt. Das gilt auch für das Qualifikationsniveau, das die Mitarbeiter der Arbeitsämter den Arbeitslosen beimessen. Fachkräfte mit Fachschulniveau und andere Fachkräfte scheinen für sich bessere Beschäftigungschancen zu sehen als Arbeitslose mit und ohne Fachkenntnisse. Wobei es auch möglich ist, dass die Vermittlungschancen dieser Arbeitslosen von den Arbeitsvermittlern im Arbeitsamt höher eingeschätzt werden, sodass sie auch eher auf Initiative der Mitarbeiter einen Gutschein bekommen. Arbeitslose, die nach Einschätzung der Arbeitsamtsmitarbeiter Berufserfahrung im angestrebten Berufsfeld haben und solche, die ihren Beruf nicht wechseln

möchten, nehmen in Ostdeutschland eher die Vermittlung durch Dritte in Anspruch als jene, die eine Jobveränderung wünschen oder über keine Berufserfahrung verfügen. Anscheinend sehen Individuen, die ihr berufsspezifisches Humankapital auch weiter im gleichen Job einsetzen möchten, für sich bessere Vermittlungschancen und nehmen deswegen das Angebot der privaten Vermittler eher in Anspruch. In den alten Bundesländern scheint dies nur für die Arbeitslosen mit Berufserfahrung im angestrebten Berufsfeld zuzutreffen, aber nicht für jene, die den Job nicht wechseln möchten.

In Ost- und Westdeutschland ist die Wahrscheinlichkeit der Maßnahmenbeteiligung bei Arbeitslosen, die eine Vermittlung in eine Teilzeitbeschäftigung wünschen geringer als bei solchen, die eine Vollzeittätigkeit anstreben. Die eingeschränkten Einsatzmöglichkeiten auf dem Arbeitsmarkt, die mit dem Wunsch nach Teilzeitarbeit verbunden sind, dürften die Arbeitslosen veranlassen, nicht auf die Vermittlung durch Dritte zurückzugreifen, da sie ihre Chancen auf dem Arbeitsmarkt insgesamt schlecht einschätzen. Dahingegen dürften die Arbeitslosen, die sogar bereit sind, einer Vollzeitbeschäftigung im Schichtdienst nachzugehen, ihre Beschäftigungschancen aufgrund ihrer Flexibilität als höher ansehen.

Die Merkmale, die den letzten Kontakt des Arbeitslosen mit dem Arbeitsmarkt beschreiben, haben in Ost- und Westdeutschland ähnliche Auswirkungen auf die Teilnahmewahrscheinlichkeit. So wirken sich die Dauer der letzten Beschäftigung negativ und die Höhe des letzten Entgeltes positiv auf die Wahrscheinlichkeit aus. Bei den Kontakten, die im Rahmen von aktiven arbeitsmarktpolitischen Maßnahmen (ABM, SAM, EGZ) zustande kamen, zeichnen sich in Ost- und Westdeutschland negative Effekte ab. Anscheinend schätzen Arbeitslose, die bereits andere arbeitsmarktpolitische Maßnahmen zur Eingliederung genutzt haben, ihre Chancen auf dem ersten Arbeitsmarkt und die Möglichkeiten der Vermittlung durch VGS negativer ein als solche, die bisher an keiner Maßnahme teilgenommen haben. Von den Variablen, die Hinweise auf die Motivation und die Verfügbarkeit der Arbeitslosen geben sollen, sind nur Abmeldungen wegen fehlender Verfügbarkeit signifikant. Auch der Leistungsbezug erhöht die Teilnahmewahrscheinlichkeit.

Wie bereits in der Diskussion der CMIA zum Ausdruck gekommen ist, hat die Erwerbsbiographie in der Schätzung einen besonderen Stellenwert.

Aus der Erwerbsbiographie soll insbesondere berücksichtigt werden, ob der Arbeitslose in den letzten Jahren in Beschäftigung oder arbeitslos war oder an Maßnahmen der Arbeitsämter teilgenommen hat. Zudem wird in der vorliegenden Schätzung berücksichtigt, wie lange der Arbeitslose in den einzelnen Zuständen während unterschiedlicher Zeitintervalle verweilte. Die Zustände der Arbeitslosigkeit und der Beschäftigung werden bis 1996 und die Teilnahme an ABM- und Fortbildungsmaßnahmen bis 2000 zurückverfolgt. Dies impliziert auch, dass die Beschäftigung erst ab 2000 um geförderte Beschäftigung bereinigt werden konnte. Ab 2000 wird also nur reguläre Beschäftigung betrachtet, davor kann diese nicht von geförderter Beschäftigung unterschieden werden. Dies führt zu einer geringfügigen Ungenauigkeit, deren Auswirkungen auf die Schätzergebnisse sich jedoch im Rahmen halten sollten. Wie bereits erwähnt, wird die nähere Vergangenheit genauer erfasst als die weiter zurückliegende. Im ersten Jahr vor der potenziellen bzw. tatsächlichen Maßnahmenteilnahme erfolgt die Betrachtung in Halbjahresschritten, danach auf Jahresbasis.

Die Schätzung zeigt, dass sich eine Beschäftigung während der letzten 5 Jahre positiv auf die Teilnahmewahrscheinlichkeit auswirkt, wobei die Wahrscheinlichkeit mit zunehmender Beschäftigungsdauer in den einzelnen Zeitabschnitten abnimmt. Bei den Arbeitslosigkeitsperioden in der Vergangenheit zeigt sich, dass weiter zurückliegende Perioden einen Einfluss auf die Teilnahmewahrscheinlichkeit haben. Die Teilnahme an ABM in Ostdeutschland hat in den zeitnahen Intervallen ebenso wie die Teilnahme an FbW einen positiven Effekt.

Die Indikatoren für den regionalen Arbeitsmarkt¹⁹ haben fast alle sowohl in Ost- als auch in Westdeutschland einen deutlichen Einfluss auf die Teilnahmewahrscheinlichkeit. So hat die regionale Arbeitslosenquote einen positiven Einfluss auf die Wahrscheinlichkeit der Maßnahmenteilnahme, während die ABM- und die FbW-Quote einen negativen Koeffizienten haben. In Westdeutschland wirkt sich die Quote der gemeldeten freien Stellen negativ auf die Teilnahmewahrscheinlichkeit aus. Anscheinend sehen hier Arbeitslose und Mitarbeiter des Arbeitsamtes bei einer relativ hohen Anzahl an gemeldeten freien Stellen keinen Bedarf, auf die Vermittlung

¹⁹ Indikatoren auf Arbeitsamtsebene.

durch VGS zurückzugreifen. Demnach dürften die privaten Vermittler eher als nützlich angesehen werden, wenn es darum geht, nicht gemeldete Stellen aufzudecken. Als letzten regionalen Indikator wurde die Quote der Arbeitslosen pro Arbeitsamtsmitarbeiter in die Schätzung aufgenommen. In Ost- und Westdeutschland hat sie einen signifikant negativen Einfluss. Eigentlich würde man vermuten, dass bei einer hohen Belastung der Mitarbeiter eher Vermittlungsgutscheine ausgeben werden, weil für die Vermittlung der Arbeitslosen nur unzureichend Zeit vorhanden ist. Die Schätzergebnisse scheinen dies nicht zu bestätigen. Möglicherweise ziehen die Agenturmitarbeiter die mögliche Arbeitseinsparung gar nicht in Betracht und sehen eher die Ausgabe des Gutscheins an sich als zusätzliche Arbeitsbelastung an, die sie lieber vermeiden (siehe dazu Bericht von sinus).

4.2 Qualität des Matchings: Balancing Tests

Es werden zwei verschiedene Varianten des Matchings durchgeführt. In der ersten Variante erfolgt ein Single Nearest Neighbour Matching mit Zurücklegen auf Basis des Propensity Scores. In der zweiten Variante werden zusätzliche Restriktionen auferlegt. Es werden den Teilnehmern nur Nichtteilnehmer zugematcht, die

- im selben regionalen Vergleichstyp arbeitslos gemeldet sind
- das selbe Geschlecht haben
- aus gleichen Altersgruppen (8) kommen (unter 25, 25 - 30, 30 - 35, ..., 50 - 55, über 55)
- bezüglich der bisherigen Arbeitslosigkeitsdauer in derselben Kategorie liegen (bis 3 Monate, 3 bis 6 Monate, 6 bis 9 Monate, 9 bis 12 Monate, 1 bis 2 Jahre, über 2 Jahre).

Neben diesen Matching-Varianten wurden weitere durchgeführt, nämlich Nearest Neighbour Matching ohne Zurücklegen und Matching mit 5 nächsten Nachbarn. Es zeigt sich jedoch, dass die geschätzten Treatment Effekte bei diesen Varianten nur vernachlässigbar geringe Abweichungen gegenüber den hier beschriebenen aufweisen. Daher wird auf die Darstellung dieser Varianten verzichtet.

Nach umfangreichen Spezifikationstests wurde eine Spezifikation für die Schätzung des Propensity Scores gefunden, die zu einer guten Ausbalancierung der Kovariaten führt. Dies gilt für beide realisierte Matching-

Varianten und sowohl für Ost- als auch für Westdeutschland, was anhand der Balancing Tests deutlich wird. Tabellen 2 und 3 zeigen für Matching-Variante 1 die Ergebnisse von zwei Tests vor und nach Matching beispielhaft für einige Kovariaten. Einerseits wird die standardisierte Differenz angegeben, andererseits der P-Wert des two-sample-T-Tests auf Abweichung in den Mittelwerten zwischen der Gruppe der Teilnehmer und derjenigen der Nichtteilnehmer.

Da das Matching nur auf Basis des Propensity Scores erfolgt, muss dieser offenbar bestmöglich ausgeglichen werden. Dies wird durch die Balancing Tests bestätigt: während die standardisierte Differenz des Propensity Scores vor dem Matching die höchste von allen betrachteten ist, liegt sie danach nahe bei null. Außerdem gibt der two-sample-T-Tests an, dass vor dem Matching hoch signifikante Unterschiede zwischen den beiden Gruppen bestehen. Nach dem Matching liegt der P-Wert des Tests nahe bei eins, es gibt also kaum noch Mittelwertdifferenzen. Generell werden die standardisierten Differenzen aller wichtigen Kovariaten durch das Matching deutlich reduziert; bei allen Kovariaten sind die standardisierten Differenzen nach Matching kleiner als zwei. Die T-Tests zeigen keine signifikanten Differenzen in den Mittelwerten nach Matching. Für Matching-Variante 2 ergeben sich ebenso gute Ergebnisse.

Außerdem wurden zwei weitere Balancing Tests durchgeführt, deren Ergebnisse ebenfalls für die gute Ausbalancierung sprechen. Beim Hotelling T^2 Test ist die Null-Hypothese, dass alle Kovariaten den gleichen Mittelwert bei den Teilnehmern und den gematchten Nichtteilnehmern haben. Diese Hypothese kann weder für Ost- noch für Westdeutschland abgelehnt werden. Bei dem von Sianesi vorgeschlagenen Test wird in einem Sample, das sich aus Teilnehmergruppe und gematchter Kontrollgruppe zusammensetzt, noch einmal eine Probitschätzung durchgeführt, bei dem die VGS-Ausgabe auf die ursprünglichen Kovariaten regressiert wird. Beim anschließenden Likelihood-ratio-Test kann die gemeinsame Insignifikanz der Kovariaten weder für Ost- noch für Westdeutschland abgelehnt werden.

**Tabelle 2: Ergebnisse der Balancing Tests für ausgewählte Kovariate, Match-
Ergebnisse für Ostdeutschland (Single Nearest Neighbor Matching
mit Zurücklegen)**

	Standardisierte Differenz		T-Test auf Unterschied in den Mittelwerten zwischen Teilnehmer- und Nichtteil- nehmergruppe (P-Wert)	
	Vor Matching	Nach Matching	Vor Matching	Nach Matching
Propensity Score	73,66	0,00	0,00	1,00
Arbeitslosigkeitsdauer bis zum (hyp.) Ausgabzeitpunkt	37,40	0,61	0,00	0,36
Alter	28,78	0,54	0,00	0,51
Ausländische Nationalität	2,10	0,34	0,00	0,66
Frauen	21,74	1,16	0,00	0,15
In Partnerschaft nicht verheiratet	1,01	0,31	0,09	0,71
allein erziehend	4,02	0,15	0,00	0,85
verheiratet	4,58	0,50	0,00	0,54
Kinder 0-3 Jahre	5,51	0,67	0,00	0,44
Kinder 3-7 Jahre	1,59	0,52	0,01	0,53
Kinder 7-14 Jahre	1,34	0,26	0,02	0,75
Behinderungsgrad (BG) 80 %	3,40	0,63	0,00	0,36
BG 50 %	6,63	0,41	0,00	0,57
BG 30 %, nicht gleichgestellt	3,02	0,65	0,00	0,39
BG 30 %, gleichgestellt	1,65	0,71	0,00	0,38
Sonstige gesundheitliche Einschränkungen	14,62	0,53	0,00	0,48
Hauptschulabschluss	10,60	0,15	0,00	0,85
Mittlere Reife	14,42	0,63	0,00	0,44
Fachhochschulreife	2,50	0,62	0,00	0,48
Abitur	2,69	1,19	0,00	0,16
betriebliche Ausbildung	9,95	1,15	0,00	0,15
Außerbetriebliche Ausbildung	8,86	0,02	0,00	0,98
Berufsfachschule, Fachschule	0,15	0,05	0,80	0,95
Universität	0,62	1,09	0,30	0,20
Fachhochschule	2,92	0,47	0,00	0,59
Spitzenkräfte	0,04	0,44	0,94	0,56
Kräfte mit Hochschulniveau	0,07	0,80	0,90	0,34
Kräfte mit Fachhochschulniveau	2,61	0,89	0,00	0,31
Kräfte mit Fachschulniveau	3,29	0,21	0,00	0,80
Fachkräfte	22,86	1,00	0,00	0,21
Sonstige Kräfte	2,08	0,93	0,00	0,25
Letzte Beschäftigungsdauer (M)	8,93	1,12	0,00	0,13
Letzte Beschäftigungsdauer (M), quadriert	9,25	0,84	0,00	0,23
Tagesentgelt bei letzter Beschäf- tigung (€)	7,72	0,93	0,00	0,11
Leistungsbezug, Höhe	24,38	1,13	0,00	0,18
Arbeitslose/Erwerbstätige	6,15	1,39	0,00	0,10
Gemeldete Stellen/Erwerbstätige	11,34	0,10	0,00	0,90
Kurzarbeiter/ Erwerbstätige	9,96	0,00	0,00	1,00
ABM/Erwerbstätige	15,38	1,16	0,00	0,15
FbW/Erwerbstätige	2,87	0,72	0,00	0,38
Arbeitslose/Mitarbeiter	3,55	0,02	0,00	0,98
Ausgabe im Juni	4,13	1,08	0,00	0,19

**Tabelle 3: Ergebnisse der Balancing Tests für ausgewählte Kovariate, Match-
Ergebnisse für Westdeutschland (Single Nearest Neighbor Mat-
ching mit Zurücklegen)**

	Standardisierte Differenz		T-Test auf Unterschied in den Mittelwerten zwischen Teilnehmer- und Nichtteil- nehmergruppe (P-Wert)	
	Vor Matching	Nach Matching	Vor Matching	Nach Matching
Propensity Score	78,87	0,00	0,00	1,00
Arbeitslosigkeitsdauer bis zum (hyp.) Ausgabezeitpunkt	27,30	0,16	0,00	0,79
Alter	28,74	1,00	0,00	0,17
Ausländische Nationalität	9,08	0,32	0,00	0,67
Frauen	7,24	0,64	0,00	0,41
In Partnerschaft nicht verheiratet	1,43	0,22	0,01	0,78
allein erziehend	2,46	0,05	0,00	0,95
verheiratet	9,48	0,01	0,00	0,99
Kinder 0-3 Jahre	4,75	0,11	0,00	0,90
Kinder 3-7 Jahre	0,05	0,01	0,93	0,99
Kinder 7-14 Jahre	0,37	1,35	0,51	0,08
Behinderungsgrad (BG) 80 %	5,36	0,67	0,00	0,31
BG 50 %	11,01	0,23	0,00	0,72
BG 30 %, nicht gleichgestellt	6,85	0,32	0,00	0,65
BG 30 %, gleichgestellt	2,24	0,26	0,00	0,73
Sonstige gesundheitliche Einschränkungen	12,12	0,73	0,00	0,33
Hauptschulabschluss	11,54	0,14	0,00	0,86
Mittlere Reife	16,99	0,65	0,00	0,43
Fachhochschulreife	7,31	0,20	0,00	0,81
Abitur	5,46	0,07	0,00	0,93
betriebliche Ausbildung	15,10	0,18	0,00	0,82
Außerbetriebliche Ausbildung	8,42	0,22	0,00	0,80
Berufsfachschule, Fachschule	3,33	0,37	0,00	0,65
Universität	2,47	0,12	0,00	0,88
Fachhochschule	4,43	0,00	0,00	1,00
Spitzenkräfte	0,44	0,42	0,40	0,55
Kräfte mit Hochschulniveau	2,36	0,15	0,00	0,86
Kräfte mit Fachhochschulniveau	4,16	0,25	0,00	0,76
Kräfte mit Fachschulniveau	4,72	0,98	0,00	0,25
Fachkräfte	17,84	0,85	0,00	0,28
Sonstige Kräfte	0,35	0,31	0,54	0,69
Letzte Beschäftigungsdauer (M)	20,45	0,12	0,00	0,86
Letzte Beschäftigungsdauer (M), quadriert	18,98	0,27	0,00	0,65
Tagesentgelt bei letzter Beschäf- tigung (€)	6,38	0,10	0,00	0,89
Leistungsbezug, Höhe	12,55	0,09	0,00	0,91
Arbeitslose/Erwerbstätige	6,67	0,21	0,00	0,79
Gemeldete Stellen/Erwerbstätige	6,11	0,34	0,00	0,65
Kurzarbeiter/ Erwerbstätige	8,51	0,10	0,00	0,89
ABM/Erwerbstätige	8,10	0,15	0,00	0,85
FbW/Erwerbstätige	14,65	0,33	0,00	0,67
Arbeitslose/Mitarbeiter	19,77	0,91	0,00	0,27
Ausgabe im Juni	2,34	0,96	0,00	0,22

Aus den Ergebnissen dieser Tests schließen wir, dass durch das Matching die beobachtbaren Merkmale von Teilnehmern und Nichtteilnehmern bei beiden Matching-Varianten sowie für Ost- und Westdeutschland gut ausgeglichen werden.

Weiterhin wurde für die Teilnehmergruppe und die gematchte Kontrollgruppe getestet, ob es genau einen Monat vor dem jeweiligen Referenzzeitpunkt (der für jede Person unterschiedlich ist) einen signifikanten Unterschied hinsichtlich des Mittelwertes der Ergebnisvariable „Verbleib in regulärer Beschäftigung“ gibt. Das gleiche wurde für monatliche Zeitabstände von bis zu 16 Monaten vor dem jeweiligen Referenzzeitpunkt untersucht. Die Ergebnisse für Matching-Variante 1 sind in den Tabellen 4 und 5 dargestellt. Es zeigt sich, dass für jeden der betrachteten Zeitabstände die Mittelwertdifferenz etwa einen Prozentpunkt oder weniger beträgt. Außerdem sind die Differenzen zumeist insignifikant. Dies spricht für eine gute Qualität des Matchings. Für Matching-Variante 2 sind die Ergebnisse ähnlich.

Tabelle 4: Verbleib in regulärer Beschäftigung 16 Monate vor (hypothetischer) Ausgabe des VGS, Teilnehmer und gematchte Kontrollgruppe Ostdeutschland

	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	t-Stat.
Anzahl	29.785	29.785			
16 Monate vor Ausgabe	26,24 %	27,30 %	-1,06 %	0,39 %	-2,7142
15 Monate vor Ausgabe	26,13 %	27,15 %	-1,02 %	0,39 %	-2,6069
14 Monate vor Ausgabe	26,21 %	26,90 %	-0,69 %	0,39 %	-1,7602
13 Monate vor Ausgabe	26,38 %	26,84 %	-0,45 %	0,39 %	-1,1586
12 Monate vor Ausgabe	26,37 %	27,06 %	-0,69 %	0,39 %	-1,7655
11 Monate vor Ausgabe	26,25 %	26,54 %	-0,29 %	0,39 %	-0,7315
10 Monate vor Ausgabe	25,27 %	25,53 %	-0,27 %	0,39 %	-0,6884
9 Monate vor Ausgabe	24,51 %	24,48 %	0,03 %	0,38 %	0,0794
8 Monate vor Ausgabe	23,37 %	23,10 %	0,28 %	0,37 %	0,7369
7 Monate vor Ausgabe	21,07 %	20,85 %	0,22 %	0,36 %	0,6247
6 Monate vor Ausgabe	17,31 %	17,61 %	-0,30 %	0,34 %	-0,8891
5 Monate vor Ausgabe	11,41 %	11,32 %	0,09 %	0,28 %	0,3229
4 Monate vor Ausgabe	5,74 %	5,64 %	0,10 %	0,20 %	0,4916
3 Monate vor Ausgabe	0,67 %	0,73 %	-0,06 %	0,07 %	-0,8157
2 Monate vor Ausgabe	0,23 %	0,34 %	-0,11 %	0,05 %	-2,2463
1 Monat vor Ausgabe	0,22 %	0,28 %	-0,06 %	0,04 %	-1,3513

Tabelle 5: Verbleib in regulärer Beschäftigung 16 Monate vor (hypothetischer) Ausgabe des VGS, Teilnehmer und gematchte Kontrollgruppe Westdeutschland

	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	t-Stat.
Anzahl	32.600	32.600			
16 Monate vor Ausgabe	39,83 %	40,12 %	-0,29 %	0,41 %	-0,71
15 Monate vor Ausgabe	39,87 %	40,27 %	-0,40 %	0,41 %	-0,96
14 Monate vor Ausgabe	38,95 %	39,30 %	-0,35 %	0,41 %	-0,86
13 Monate vor Ausgabe	37,79 %	38,66 %	-0,87 %	0,41 %	-2,11
12 Monate vor Ausgabe	37,20 %	38,06 %	-0,86 %	0,41 %	-2,10
11 Monate vor Ausgabe	35,66 %	36,35 %	-0,68 %	0,40 %	-1,69
10 Monate vor Ausgabe	33,74 %	34,10 %	-0,37 %	0,40 %	-0,92
9 Monate vor Ausgabe	31,42 %	31,73 %	-0,31 %	0,39 %	-0,78
8 Monate vor Ausgabe	28,61 %	28,69 %	-0,08 %	0,38 %	-0,21
7 Monate vor Ausgabe	24,90 %	24,78 %	0,11 %	0,36 %	0,31
6 Monate vor Ausgabe	20,26 %	20,94 %	-0,69 %	0,34 %	-2,02
5 Monate vor Ausgabe	13,56 %	13,62 %	-0,06 %	0,29 %	-0,21
4 Monate vor Ausgabe	6,77 %	6,52 %	0,25 %	0,21 %	1,17
3 Monate vor Ausgabe	1,06 %	1,22 %	-0,15 %	0,09 %	-1,71
2 Monate vor Ausgabe	0,50 %	0,64 %	-0,14 %	0,06 %	-2,26
1 Monat vor Ausgabe	0,39 %	0,53 %	-0,15 %	0,06 %	-2,56

4.3 Schätzung der Average Treatment Effekte für die Teilnehmer

Tabelle 6 zeigt die geschätzten Average Treatment Effekte für die Teilnehmer in Ostdeutschland nach Variante 1. Die geschätzten Effekte für Variante 2 unterscheiden sich nicht nennenswert davon. Die Ergebnisse sind also in dieser Hinsicht robust. Es wird zunächst der Verbleib in Beschäftigung betrachtet, d. h. ob eine Person sich zu einem bestimmten Zeitpunkt in Beschäftigung *befindet*. In diese Größe fließen einerseits die Übergänge in Beschäftigung ein, andererseits aber auch die nachfolgenden Abgänge aus der Beschäftigung. Im zweiten Schritt sollen die Größen Übergänge in Beschäftigung und Dauer der Beschäftigung getrennt betrachtet werden.

Einen Monat nach dem Analysezeitpunkt befinden sich rund 8,1 Prozent der VGS-Bezieher in regulärer Beschäftigung, aber nur 4,8 Prozent der zugematchten Nichtteilnehmer. Der geschätzte ATT beträgt somit rund 3,3 Prozentpunkte. Er steigt bis zum dritten Monat auf 4,8 Prozentpunkte an. In den darauf folgenden 3 Monaten bleibt er ungefähr konstant, da sich die Anteile in Beschäftigung in den beiden Gruppen dann nahezu parallel entwickeln. In der rechten Spalte ist der Anteil der eingelösten VGS an

den ausgegebenen nach sechs Zeitintervallen angegeben. Dabei wird nur berücksichtigt, ob der VGS bis zu diesem Zeitpunkt eingelöst wurde, aber nicht, ob der Arbeitslose noch in der Beschäftigung ist. Bei einem Vergleich mit der Spalte „Verbleib in regulärer Beschäftigung“ für die Teilnehmer ist somit zu beachten, dass einerseits VGS-Vermittelte schon wieder arbeitslos geworden sein können, andererseits aber auch viele Teilnehmer auf anderem Wege in Beschäftigung gelangen. Der Anteil der Einlösungen liegt einen Monat nach der Ausgabe bei 6,3 Prozent und steigt bis auf 12,2 Prozent nach 6 Monaten an.

Tabelle 6: Geschätzte Average Treatment Effekte für die Teilnehmer in Ostdeutschland: Verbleib in sozialversicherungspflichtiger Beschäftigung laut BSt und Verbleib in Arbeitslosigkeit laut BewA; für Teilnehmer Anteil der ausgegebenen VGS, die eingelöst werden

Dauer nach Analysezeitpunkt (Monate)	<i>Verbleib in regulärer Beschäftigung</i>				<i>Für Teilnehmer: Anteil der Einlösungen</i>
	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
Anzahl	29.785	29.785			
1	8,09 %	4,76 %	3,33 %	0,21 %	6,33 %
2	12,50 %	8,14 %	4,36 %	0,26 %	9,07 %
3	15,06 %	10,23 %	4,83 %	0,29 %	10,61 %
4	16,68 %	11,75 %	4,93 %	0,31 %	11,26 %
5	17,38 %	12,52 %	4,86 %	0,31 %	11,86 %
6	17,17 %	12,37 %	4,80 %	0,31 %	12,24 %

Dauer nach Analysezeitpunkt (Monate)	<i>Verbleib in Arbeitslosigkeit</i>			
	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler
1	84,08 %	88,00 %	-3,92 %	0,30 %
2	74,03 %	78,79 %	-4,76 %	0,37 %
3	66,89 %	70,90 %	-4,01 %	0,41 %
4	61,40 %	64,44 %	-3,03 %	0,43 %
5	57,51 %	60,17 %	-2,66 %	0,44 %
6	55,91 %	58,17 %	-2,26 %	0,44 %

Aus den Ergebnissen lassen sich die Mitnahmeeffekte direkt ablesen. Mitnahmeeffekte im ökonomischen Sinne werden gerade durch die Höhe des Verbleibs in Beschäftigung in der Kontrollgruppe abgeschätzt, also nach 6 Monaten 12,4 Prozent. Schädlich sind diese Mitnahmeeffekte aber vor allem insoweit, als sie mit Auszahlungen von Prämien verbunden sind und damit besonders hohe zusätzliche Kosten erzeugen. Allerdings führt auch schon die Ausgabe der VGS zu Verwaltungsaufwand. Die Kostenseite wird in Kapitel 6 betrachtet.

Betrachtet man den Verbleib in Arbeitslosigkeit, so ergibt sich ein ähnliches Bild. Einen Monat nach dem Analysedatum sind von den Teilnehmern rund 3,9 Prozentpunkte weniger arbeitslos als von den zugematchten Nichtteilnehmern. Dieser Abstand steigt nach zwei Monaten auf 4,8 Prozentpunkte und sinkt dann bis zum sechsten Monat wieder auf 2,3 Prozentpunkte. Es ist festzustellen, dass ein wachsender Teil der Personen sich weder in Beschäftigung noch in Arbeitslosigkeit befindet. Viele von diesen Personen gehen in ABM oder geförderte Beschäftigung über. Von den VGS-Beziehern sind dies nach 6 Monaten 13,8 Prozent, von den Nichtteilnehmern 12,9 Prozent.

Tabelle 7: Geschätzte Average Treatment Effekte für die Teilnehmer in Westdeutschland: Verbleib in sozialversicherungspflichtiger Beschäftigung laut BSt und Verbleib in Arbeitslosigkeit laut BewA; für Teilnehmer Anteil der ausgegebenen VGS, die eingelöst werden

Dauer nach Analysezeitpunkt (Monate)	Verbleib in regulärer Beschäftigung			Std.fehler	Für Teilnehmer: Anteil der Einlösungen
	Teilnehmer	Kontrollgruppe	Differenz		
Anzahl	32.600	32.600			
1	6,52 %	4,78 %	1,74 %	0,19 %	3,37 %
2	11,18 %	8,19 %	2,99 %	0,25 %	4,85 %
3	14,28 %	10,72 %	3,56 %	0,28 %	5,77 %
4	16,10 %	12,50 %	3,60 %	0,29 %	6,20 %
5	17,17 %	13,58 %	3,59 %	0,30 %	6,51 %
6	17,66 %	13,94 %	3,72 %	0,30 %	6,75 %

Dauer nach Analysezeitpunkt (Monate)	Verbleib in Arbeitslosigkeit			Std.fehler
	Teilnehmer	Kontrollgruppe	Differenz	
1	88,60 %	89,49 %	-0,90 %	0,26 %
2	79,53 %	80,66 %	-1,13 %	0,34 %
3	71,97 %	73,29 %	-1,32 %	0,38 %
4	66,26 %	67,17 %	-0,90 %	0,40 %
5	61,72 %	62,50 %	-0,78 %	0,41 %
6	59,14 %	59,78 %	-0,64 %	0,41 %

In Tabelle 7 sind die entsprechenden Ergebnisse der Variante 1 für Westdeutschland dargestellt (auch hier gibt es keine nennenswerten Unterschiede zwischen den beiden Varianten). Der geschätzte durchschnittliche Effekt für die Teilnehmer steigt von 1,7 Prozentpunkten nach einem Monat auf 3,6 Prozentpunkte nach drei Monaten, danach bleibt der Effekt etwa gleich. Er ist somit nach allen betrachteten Zeitintervallen deutlich kleiner als in Ostdeutschland (gut einen Prozentpunkt). Der Anteil der Einlösungen

steigt hier von 3,4 Prozent nach einem auf 6,7 Prozent nach 6 Monaten und ist damit sehr viel kleiner als im Osten. Dies wird sich in der Kosten-Nutzen-Analyse positiv widerspiegeln.

Der Unterschied zu Ostdeutschland ist größer, wenn man den Verbleib in Arbeitslosigkeit laut BewA betrachtet. Der Effekt beträgt hier nur bis zu 1,3 Prozentpunkte (nach 3 Monaten) zugunsten der Teilnehmer. Dies liegt daran, dass in Westdeutschland mehr Arbeitslose in andere Zustände abgehen, was insbesondere für die Nichtteilnehmer gilt.

4.4 Gruppenspezifische Treatment Effekte

Es werden Average Treatment Effekte für unterschiedliche Subsamples angegeben. Dazu werden die Ergebnisse der Matching-Variante 2 verwendet, da bei dieser gewährleistet ist, dass das Matching nur innerhalb von bestimmten Subsamples erfolgt. So werden bspw. nicht Männer mit Frauen gematcht. Auch die Einlösungsquoten für alle Untergruppen werden dargestellt. Bei der Betrachtung der Zahlen ist zu beachten, dass der Erfolg der Gutscheine am geschätzten Treatment Effekt zu messen ist. Bei gegebenem Treatment Effekt stellen höhere Einlösungsquoten hingegen höhere Kosten durch mehr Prämienzahlungen dar.

In Tabelle 8 und 9 werden die geschätzten Effekte nach der Art des VGS differenziert. Es ist zu erkennen, dass die Effekte auf den Verbleib in regulärer Beschäftigung mit der Höhe der Wertstufe abnehmen. Bei den Gutscheinen über 1.500 € beträgt der geschätzte Effekt in Ostdeutschland nach 6 Monaten 6,5 Prozentpunkte, bei der mittleren Wertstufe 5,3 und bei der höchsten Stufe 3,6 Prozentpunkte. Dies ist dadurch erklärbar, dass die Personen mit geringerwertigem VGS kürzer arbeitslos sind und somit durch die privaten Vermittler besser vermittelt werden können. Dies wird offenbar nicht dadurch kompensiert, dass die privaten Vermittler bei Arbeitslosen mit längerer Arbeitslosigkeitsdauer wegen der höheren Prämie auch einen größeren Anreiz haben, sich um die Vermittlung zu bemühen.

Auch in Westdeutschland sind sowohl die geschätzten Treatment Effekte als auch der Anteil der Einlösungen für die höherwertigen VGS geringer.

Tabelle 8: Geschätzte Average Treatment Effekte differenziert nach Art des Gutscheins in Ostdeutschland

	<i>Verbleib in regulärer Beschäftigung</i>				<i>Für Teilnehmer: Anteil der Einlösungen</i>
VGS über 1.500 € (9.416)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	12,59 %	8,01 %	4,58 %	0,47 %	8,53 %
2	19,67 %	13,81 %	5,86 %	0,58 %	12,17 %
3	23,86 %	17,25 %	6,62 %	0,63 %	14,07 %
4	26,13 %	19,03 %	7,09 %	0,65 %	14,76 %
5	27,20 %	20,40 %	6,80 %	0,66 %	15,57 %
6	26,85 %	20,33 %	6,52 %	0,66 %	16,00 %
VGS über 2.000 € (5.460)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	9,27 %	5,42 %	3,85 %	0,57 %	7,27 %
2	14,73 %	9,82 %	4,91 %	0,72 %	10,53 %
3	17,80 %	12,49 %	5,31 %	0,79 %	12,69 %
4	19,65 %	14,43 %	5,22 %	0,83 %	13,39 %
5	20,44 %	14,80 %	5,64 %	0,84 %	13,96 %
6	19,36 %	14,08 %	5,27 %	0,83 %	14,30 %
VGS über 2.500 € (14.909)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	4,82 %	2,35 %	2,47 %	0,22 %	4,60 %
2	7,16 %	3,98 %	3,17 %	0,28 %	6,58 %
3	8,50 %	5,11 %	3,39 %	0,30 %	7,67 %
4	9,63 %	5,92 %	3,71 %	0,32 %	8,26 %
5	10,05 %	6,53 %	3,53 %	0,33 %	8,74 %
6	10,26 %	6,65 %	3,61 %	0,34 %	9,11 %

Tabelle 9: Geschätzte Average Treatment Effekte differenziert nach Art des Gutscheins in Westdeutschland

	<i>Verbleib in regulärer Beschäftigung</i>				<i>Für Teilnehmer: Anteil der Einlösungen</i>
VGS über 1.500 € (10.722)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	8,74 %	7,13 %	1,60 %	0,39 %	3,70 %
2	15,52 %	12,05 %	3,47 %	0,50 %	5,40 %
3	20,15 %	15,20 %	4,94 %	0,55 %	6,58 %
4	22,64 %	17,35 %	5,29 %	0,58 %	7,03 %
5	23,88 %	18,40 %	5,47 %	0,59 %	7,45 %
6	24,96 %	18,58 %	6,38 %	0,59 %	7,74 %
VGS über 2.000 € (6.519)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	8,30 %	5,80 %	2,50 %	0,51 %	4,39 %
2	13,68 %	9,37 %	4,31 %	0,63 %	5,92 %
3	17,04 %	12,20 %	4,85 %	0,70 %	6,87 %
4	18,96 %	14,02 %	4,94 %	0,74 %	7,47 %
5	20,34 %	15,11 %	5,23 %	0,76 %	7,72 %
6	20,23 %	15,26 %	4,97 %	0,76 %	7,96 %
VGS über 2.500 € (15.359)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	4,22 %	2,67 %	1,55 %	0,22 %	2,72 %
2	7,10 %	4,67 %	2,42 %	0,29 %	4,01 %
3	9,02 %	6,26 %	2,76 %	0,33 %	4,73 %
4	10,33 %	7,47 %	2,86 %	0,35 %	5,08 %
5	11,15 %	8,07 %	3,08 %	0,36 %	5,34 %
6	11,47 %	8,51 %	2,96 %	0,37 %	5,53 %

In den Tabellen 10 und 11 werden die geschätzten Effekte für Männer und Frauen vergleichend dargestellt. Zunächst wird deutlich, dass sowohl im Osten als auch im Westen bei den Männern nach jedem der betrachteten Zeitintervalle ein größerer Anteil der VGS-Bezieher und der Nichtteilnehmer in Beschäftigung übergegangen ist als bei den Frauen. Diese Unterschiede sind im Osten sehr deutlich, im Westen nur gering.

Die geschätzten Treatment Effekte für Ostdeutschland sind vor allem in den ersten drei Monaten nach dem Analysedatum bei den Männern deutlich höher. Sie steigen bei diesen von 4,7 Prozentpunkten nach einem Monat auf 6,6 Prozentpunkte nach drei Monaten, bei den Frauen hingegen

nur von 1,3 auf 1,7 Prozentpunkte. Danach sinkt der geschätzte ATT für die Männer jedoch wieder bis auf 6,0 Prozentpunkte nach 6 Monaten, bei den Frauen steigt er hingegen weiter bis auf 2,9 Prozentpunkte. Für Westdeutschland sind die Unterschiede im ATT sehr viel geringer, außerdem bleiben sie im Zeitablauf zumeist konstant unter einem Prozentpunkt.

In Ostdeutschland beträgt der Anteil der Einlösungen nach 6 Monaten bei den Männern über 15 Prozent der ausgegebenen VGS, während er bei den Frauen nur knapp halb so groß ist. Auch für diese Größe sind die Unterschiede in Westdeutschland deutlich geringer.

Tabelle 10: Geschätzte Average Treatment Effekte differenziert nach Geschlecht in Ostdeutschland

	<i>Verbleib in regulärer Beschäftigung</i>				<i>Für Teilnehmer: Anteil der Einlösungen</i>
Männer (18.429)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	10,67 %	5,99 %	4,68 %	0,35 %	8,05%
2	16,36 %	10,32 %	6,04 %	0,43 %	11,46%
3	19,51 %	12,87 %	6,64 %	0,47 %	13,44%
4	20,98 %	14,30 %	6,68 %	0,49 %	14,19%
5	21,49 %	15,05 %	6,44 %	0,50 %	14,92%
6	20,55 %	14,51 %	6,04 %	0,49 %	15,32%
Frauen (11.356)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	3,90 %	2,62 %	1,29 %	0,34 %	3,54 %
2	6,23 %	4,66 %	1,58 %	0,44 %	5,20 %
3	7,84 %	6,13 %	1,71 %	0,50 %	6,03 %
4	9,70 %	7,27 %	2,42 %	0,54 %	6,50 %
5	10,70 %	8,17 %	2,53 %	0,57 %	6,88 %
6	11,68 %	8,81 %	2,87 %	0,59 %	7,24 %

Tabelle 11: Geschätzte Average Treatment Effekte differenziert nach Geschlecht in Westdeutschland

	<i>Verbleib in regulärer Beschäftigung</i>				<i>Für Teilnehmer: Anteil der Einlösungen</i>
Männer (21.623)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	7,25%	5,13%	2,12%	0,28%	3,64%
2	12,13%	8,70%	3,43%	0,36%	5,28%
3	15,39%	11,09%	4,30%	0,40%	6,27%
4	16,89%	12,62%	4,27%	0,42%	6,64%
5	17,80%	13,28%	4,51%	0,42%	6,98%
6	17,98%	13,36%	4,62%	0,43%	7,24%
Frauen (10.977)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	5,08%	4,04%	1,05%	0,44%	2,84%
2	9,32%	6,74%	2,58%	0,56%	4,00%
3	12,10%	9,00%	3,10%	0,64%	4,78%
4	14,55%	10,87%	3,68%	0,69%	5,34%
5	15,95%	12,08%	3,87%	0,72%	5,58%
6	17,02%	12,81%	4,21%	0,74%	5,78%

In den Tabellen 12 und 13 sind die geschätzten Effekte nach drei Altersgruppen differenziert dargestellt. Offenbar werden die durchschnittlichen Treatment Effekte auf den Verbleib in regulärer Beschäftigung mit zunehmendem Alter kleiner. Während sie bei den unter 25-Jährigen von 5,2 Prozentpunkten nach einem Monat auf 7,8 Prozentpunkte nach 6 Monaten steigen, sind es bei den über 50-Jährigen nur 2,6 und 3,9 Prozentpunkte. In der mittleren Altersgruppe liegen die Effekte dazwischen. Alle diese Schätzungen sind statistisch signifikant auf dem 5 %-Niveau. Außerdem wird deutlich, dass bei den jüngeren Arbeitslosen auch ein größerer Teil der Gutscheine eingelöst wird: 15,7 Prozent nach 6 Monaten bei den unter 25-Jährigen und nur 10,4 Prozent bei den über 50-Jährigen.

Tabelle 12: Geschätzte Average Treatment Effekte differenziert nach Altersgruppen in Ostdeutschland

	<i>Verbleib in regulärer Beschäftigung</i>				<i>Für Teilnehmer: Anteil der Einlösungen</i>
Unter 25 Jahre (3.468)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	13,58 %	8,36 %	5,22 %	1,72 %	8,94 %
2	20,73 %	14,33 %	6,40 %	2,16 %	12,54 %
3	25,29 %	18,25 %	7,04 %	2,37 %	14,24 %
4	28,23 %	20,30 %	7,93 %	2,47 %	14,73 %
5	29,35 %	21,91 %	7,44 %	2,53 %	15,31 %
6	29,38 %	21,54 %	7,84 %	2,52 %	15,69 %
25 bis 49 Jahre (21.583)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	7,89 %	4,61 %	3,27 %	0,27 %	6,13 %
2	12,18 %	7,95 %	4,23 %	0,34 %	8,81 %
3	14,71 %	10,13 %	4,57 %	0,37 %	10,39 %
4	16,32 %	11,42 %	4,90 %	0,39 %	11,06 %
5	17,03 %	12,25 %	4,78 %	0,40 %	11,69 %
6	16,77 %	12,20 %	4,56 %	0,40 %	12,10 %
50 Jahre und älter (4.734)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	5,01 %	2,43 %	2,58 %	0,73 %	5,34 %
2	7,92 %	4,61 %	3,32 %	0,98 %	7,71 %
3	9,19 %	5,24 %	3,95 %	1,04 %	8,96 %
4	9,84 %	6,19 %	3,65 %	1,12 %	9,61 %
5	10,20 %	6,32 %	3,89 %	1,13 %	10,05 %
6	10,05 %	6,19 %	3,87 %	1,12 %	10,35 %

In Westdeutschland sind die Unterschiede zwischen den Altersgruppen noch größer. Für die unter 25-Jährigen beträgt der geschätzte Effekt nach 6 Monaten 9,8 Prozentpunkte, für die über 50-Jährigen nur 2,3 Prozentpunkte. Auch der Anteil der Einlösungen variiert stärker (von 10,5 bis 4,5 Prozent). Hier wird besonders deutlich, dass die Höhe der Mitnahmeeffekte im ökonomischen Sinne noch wenig über die Kosten-Nutzen-Relation aussagt. Bei der jüngsten Altersgruppe sind die Mitnahmeeffekte mit 21,8 Prozent zwar viel höher als bei der ältesten mit 5,5 Prozent. Jedoch scheint bei den jüngeren ein sehr viel geringerer Teil der Mitnahmen mit Prämienauszahlungen verbunden zu sein, da die Einlösungen den geschätzten Beschäftigungseffekt nur geringfügig übersteigen. Bei den älte-

ren übersteigen die Einlösungen den Effekt nach 6 Monaten fast um das Doppelte. Dies deutet auf eine bessere Kosten-Nutzen-Relation bei den jüngeren Arbeitslosen hin.

Tabelle 13: Geschätzte Average Treatment Effekte differenziert nach Altersgruppen in Westdeutschland

	<i>Verbleib in regulärer Beschäftigung</i>				<i>Für Teilnehmer: Anteil der Einlösungen</i>
Unter 25 Jahre (3.139)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	14,05 %	9,02 %	5,03 %	2,04 %	6,31 %
2	22,08 %	14,11 %	7,96 %	2,48 %	8,57 %
3	27,11 %	17,81 %	9,30 %	2,72 %	9,59 %
4	28,70 %	21,38 %	7,33 %	2,90 %	9,94 %
5	30,17 %	22,17 %	8,00 %	2,94 %	10,26 %
6	31,54 %	21,76 %	9,78 %	2,93 %	10,48 %
25 bis 49 Jahre (24.733)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	6,23 %	4,80 %	1,44 %	0,24 %	3,28 %
2	10,90 %	8,16 %	2,75 %	0,31 %	4,75 %
3	14,07 %	10,58 %	3,49 %	0,35 %	5,70 %
4	16,15 %	12,18 %	3,97 %	0,37 %	6,15 %
5	17,31 %	13,12 %	4,19 %	0,38 %	6,45 %
6	17,77 %	13,54 %	4,23 %	0,38 %	6,71 %
50 Jahre und älter (4.728)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	3,02 %	1,76 %	1,27 %	0,64 %	1,92 %
2	5,41 %	3,41 %	2,01 %	0,88 %	2,88 %
3	6,85 %	4,44 %	2,41 %	1,00 %	3,60 %
4	7,47 %	5,01 %	2,45 %	1,06 %	3,96 %
5	7,85 %	5,44 %	2,41 %	1,10 %	4,31 %
6	7,87 %	5,54 %	2,33 %	1,10 %	4,46 %

Weiterhin werden die geschätzten Effekte nach regionalen Vergleichstypen differenziert (Tabellen 14 und 15). Für Ostdeutschland zeigt sich, dass sich die geschätzten Effekte in den regionalen Vergleichstypen wenig unterscheiden. Allein in Typ Ic ist der Effekt mit 3,9 Prozentpunkten etwas geringer als in den anderen Typen mit jeweils 5 Prozentpunkten. Der Anteil der Einlösungen ist hingegen umso höher, je schlechter die Arbeitsmarktlage in der Region ist. So werden im Vergleichstyp Ia (Bezirke in Ostdeutschland mit schlechtesten Arbeitsmarktbedingungen) nach 6 Monaten 13,8 Prozent der VGS eingelöst.

Tabelle 14: Geschätzte Average Treatment Effekte differenziert nach regionalen Vergleichstypen in Ostdeutschland

	<i>Verbleib in regulärer Beschäftigung</i>				<i>Für Teilnehmer: Anteil der Einlösungen</i>
Vergleichstyp Ia (3.560)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	8,29 %	4,19 %	4,10 %	1,23%	6,85 %
2	12,13 %	7,53 %	4,61 %	1,60%	10,06 %
3	13,88 %	9,30 %	4,58 %	1,75%	11,97 %
4	15,45 %	10,53 %	4,92 %	1,85%	12,70 %
5	16,21 %	11,04 %	5,17 %	1,89%	13,29 %
6	15,84 %	10,84 %	5,00 %	1,87%	13,79 %
Vergleichstyp Ib (19.913)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	8,15 %	4,76 %	3,39 %	0,29 %	6,41 %
2	12,74 %	8,18 %	4,55 %	0,37 %	9,20 %
3	15,36 %	10,34 %	5,02 %	0,40 %	10,69 %
4	17,05 %	11,63 %	5,42 %	0,42 %	11,35 %
5	17,60 %	12,44 %	5,16 %	0,43 %	11,92 %
6	17,31 %	12,29 %	5,02 %	0,43 %	12,35 %
Vergleichstyp Ic (5.025)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	8,36 %	4,98 %	3,38 %	0,96 %	6,35 %
2	12,96 %	9,09 %	3,86 %	1,25 %	8,94 %
3	15,70 %	11,52 %	4,18 %	1,38 %	10,33 %
4	17,05 %	13,03 %	4,02 %	1,46 %	10,83 %
5	18,17 %	14,11 %	4,06 %	1,50 %	11,58 %
6	18,17 %	14,23 %	3,94 %	1,51 %	11,74 %
Vergleichstyp IIa (1.287)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	5,59 %	4,20 %	1,40 %	3,22 %	3,57 %
2	8,08 %	5,91 %	2,18 %	3,78 %	4,90 %
3	11,19 %	7,69 %	3,50 %	4,28 %	6,76 %
4	12,90 %	9,01 %	3,89 %	4,60 %	7,46 %
5	14,14 %	9,56 %	4,58 %	4,72 %	7,93 %
6	14,76 %	9,79 %	4,97 %	4,78 %	8,24 %

In Westdeutschland variieren die geschätzten durchschnittlichen Treatment Effekte auf die Teilnehmer nach 6 Monaten deutlich zwischen den Vergleichstypen, von 6,9 Prozentpunkten für Typ IIa (ohne Dresden) bis 1,4 Prozentpunkt in Typ Vc. In den Vergleichstypen IIIa, IIIb, IV, Va, Vb und Vc sind die geschätzten Effekte nach 6 Monaten zwar positiv, aber nicht signifikant. Bei Typ Vc liegt das vor allem an dem niedrigen geschätzten Effekt; bei den anderen Typen ist dies eher auf die geringe Zahl von ausgegebenen VGS zurückzuführen.

Auch die Quote der VGS-Einlösungen nach 6 Monaten variiert deutlich zwischen den regionalen Vergleichstypen, zwischen 10,6 Prozent in Typ IIIb und 4,5 Prozent in Typ IIb. Betrachtet man das Verhältnis zwischen (Nutzen stiftendem) Treatment Effekt und (Kosten treibenden) Einlösungen, so ist dieses in Vergleichstyp IIa und IV besonders günstig, in Typ Vc äußerst ungünstig.

Tabelle 15 Geschätzte Average Treatment Effekte differenziert nach regionalen Vergleichstypen in Westdeutschland

		<i>Verbleib in regulärer Beschäftigung</i>			<i>Für Teilnehmer: Anteil der Einlösungen</i>
Vergleichstyp IIa (5.759)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	6,79 %	3,59 %	3,20 %	0,77 %	4,20 %
2	11,41 %	6,32 %	5,09 %	0,99 %	6,20 %
3	14,95 %	8,37 %	6,58 %	1,13 %	7,22 %
4	16,25 %	9,65 %	6,60 %	1,20 %	7,76 %
5	16,93 %	10,31 %	6,62 %	1,23 %	8,16 %
6	17,42 %	10,49 %	6,93 %	1,24 %	8,42 %
Vergleichstyp IIb (8.335)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	5,29 %	4,13 %	1,16 %	0,57 %	2,17 %
2	9,29 %	6,81 %	2,47 %	0,72 %	3,19 %
3	12,42 %	9,18 %	3,24 %	0,82 %	3,78 %
4	14,06 %	10,81 %	3,25 %	0,88 %	4,12 %
5	15,13 %	11,66 %	3,47 %	0,91 %	4,33 %
6	15,64 %	12,19 %	3,46 %	0,93 %	4,52 %
Vergleichstyp IIIa (3.098)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	5,87 %	4,68 %	1,19 %	1,52 %	3,07 %
2	9,39 %	8,13 %	1,26 %	1,96 %	4,78 %
3	12,20 %	10,30 %	1,90 %	2,18 %	5,62 %
4	13,78 %	11,81 %	1,97 %	2,31 %	5,91 %
5	14,36 %	12,49 %	1,87 %	2,36 %	6,23 %
6	14,78 %	12,17 %	2,61 %	2,34 %	6,58 %
Vergleichstyp IIIb (2.055)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	9,05 %	5,89 %	3,16 %	2,53 %	5,94 %
2	14,16 %	9,88 %	4,28 %	3,19 %	7,93 %
3	17,23 %	12,55 %	4,67 %	3,54 %	9,00 %
4	19,51 %	14,06 %	5,45 %	3,72 %	9,78 %
5	21,22 %	15,28 %	5,94 %	3,85 %	10,17 %
6	21,36 %	15,33 %	6,03 %	3,85 %	10,56 %

Vergleichstyp IIIc (5.426)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	7,00 %	5,33 %	1,68 %	0,95 %	3,93 %
2	12,27 %	8,99 %	3,28 %	1,21 %	5,33 %
3	15,39 %	11,46 %	3,93 %	1,35 %	6,43 %
4	16,90 %	13,01 %	3,89 %	1,42 %	6,86 %
5	18,17 %	13,51 %	4,66 %	1,44 %	7,13 %
6	18,47 %	13,80 %	4,66 %	1,46 %	7,35 %
Vergleichstyp IV (2.695)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	6,57 %	4,68 %	1,89 %	1,74 %	3,01 %
2	11,13 %	7,46 %	3,67 %	2,17 %	4,45 %
3	14,29 %	10,43 %	3,86 %	2,51 %	5,38 %
4	16,55 %	12,24 %	4,30 %	2,69 %	5,83 %
5	17,92 %	13,06 %	4,86 %	2,77 %	6,16 %
6	18,48 %	13,32 %	5,16 %	2,79 %	6,35 %
Vergleichstyp Va (649)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	10,32 %	7,40 %	2,93 %	8,70 %	5,39 %
2	15,72 %	12,48 %	3,24 %	10,97 %	7,09 %
3	18,80 %	14,64 %	4,16 %	11,74 %	8,63 %
4	20,80 %	17,57 %	3,24 %	12,63 %	9,09 %
5	19,72 %	17,87 %	1,85 %	12,71 %	9,55 %
6	21,73 %	19,11 %	2,62 %	13,04 %	9,86 %
Vergleichstyp Vb (3.410)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	6,89 %	5,45 %	1,44 %	1,49 %	2,73 %
2	12,84 %	9,53 %	3,31 %	1,93 %	4,13 %
3	15,40 %	11,79 %	3,61 %	2,11 %	5,22 %
4	18,04 %	14,05 %	3,99 %	2,27 %	5,72 %
5	19,79 %	15,48 %	4,31 %	2,37 %	6,01 %
6	20,29 %	16,04 %	4,25 %	2,40 %	6,25 %
Vergleichstyp Vc (1.173)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	5,71 %	7,42 %	-1,71 %	4,82 %	3,24 %
2	10,83 %	11,85 %	-1,02 %	5,95 %	4,35 %
3	13,73 %	13,81 %	-0,09 %	6,36 %	5,29 %
4	17,05 %	15,35 %	1,71 %	6,65 %	5,46 %
5	17,90 %	17,22 %	0,68 %	6,96 %	5,88 %
6	18,67 %	17,31 %	1,36 %	6,98 %	5,88 %

Aggregiert man die Vergleichstypen zu regionalen Strategietypen (Tabelle 16) reduziert sich die Variation bei den geschätzten Effekte. Sie liegen zwischen 3,4 Prozentpunkten für Typ V und 5,2 Prozentpunkten für Typ

IV. Im Typ III beträgt die Einlösungsquote 7,8 Prozent, der geschätzte Effekt aber nur 4,3 Prozent.

Tabelle 16: Geschätzte Average Treatment Effekte differenziert nach regionalen Strategietypen in Westdeutschland

	<i>Verbleib in regulärer Beschäftigung</i>				<i>Für Teilnehmer: Anteil der Einlösungen</i>
Strategietyp II (14.094)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	5,90 %	3,91 %	1,99 %	0,35 %	3,00 %
2	10,15 %	6,61 %	3,54 %	0,45 %	4,42 %
3	13,45 %	8,85 %	4,60 %	0,52 %	5,19 %
4	14,96 %	10,34 %	4,62 %	0,55 %	5,61 %
5	15,86 %	11,11 %	4,75 %	0,57 %	5,90 %
6	16,37 %	11,49 %	4,87 %	0,57 %	6,12 %
Strategietyp III (10.579)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	7,07 %	5,25 %	1,82 %	0,51 %	4,06 %
2	11,80 %	8,91 %	2,88 %	0,65 %	5,67 %
3	14,81 %	11,33 %	3,48 %	0,73 %	6,69 %
4	16,49 %	12,87 %	3,63 %	0,77 %	7,15 %
5	17,65 %	13,56 %	4,09 %	0,78 %	7,46 %
6	17,95 %	13,62 %	4,33 %	0,79 %	7,75 %
Strategietyp IV entspricht Vergleichstyp IV					
Strategietyp V (5.232)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
1	7,05 %	6,14 %	0,92 %	1,04 %	3,17 %
2	12,75 %	10,42 %	2,33 %	1,33 %	4,55 %
3	15,44 %	12,60 %	2,85 %	1,44 %	5,66 %
4	18,16 %	14,77 %	3,38 %	1,54 %	6,08 %
5	19,36 %	16,17 %	3,19 %	1,60 %	6,42 %
6	20,11 %	16,70 %	3,40 %	1,62 %	6,61 %

Als nächstes werden die Treatment Effekte differenziert nach den vom IAW generierten Typen. Dabei wurde die Typisierung nach Stellenwert/Informiertheit einerseits und Intensität andererseits verwendet (Variante b4 in Kapitel 2). Der Faktor „Stellenwert/Informiertheit“ drückt aus, wie gründlich und wie überzeugt die Agenturen die Begleitung des VGS durchführen. Die Informiertheit wird daran festgemacht, wie gut die Mitarbeiter

über (mögliche) Missbrauchsfälle informiert sind. Demgegenüber bildet der Faktor „Intensität“ den Personaleinsatz zur Durchführung von VGS und die Ausgabepraxis von VGS an die Anspruchsberechtigten ab.

In Tabelle 17 wird deutlich, dass in Ostdeutschland ein höherer Stellenwert bzw. eine bessere Informiertheit mit einem höheren Treatment Effekt bei gleich bleibenden Einlösungsquoten verbunden sind. Dies könnte darauf hindeuten, dass es diesen Ämtern durch mehr Kontrolle gelingt, das Instrument effizienter einzusetzen. Eine höhere Intensität ist mit deutlich höheren Ausgabequoten verbunden, aber mit niedrigeren Treatment Effekten und Einlösungsquoten. Detailliertere Analysen zeigen, dass dabei vor allem die Ausgabeintensität eine Rolle spielt. Das Ergebnis könnte auf abnehmende Grenzerträge des Instrumentes hinweisen. Eine forcierte Ausgabe führt demnach dazu, dass zunehmend auch solche Arbeitslose einen VGS erhalten, für die das Instrument weniger geeignet ist. Außerdem dürften sich die VGS-Bezieher auf dem Arbeitsmarkt zunehmend gegenseitig Konkurrenz machen, so dass die Einlösungsquote sinkt.

Tabelle 17: Regionaler IAW-Typ in Ostdeutschland: Ausgabequote, Treatment Effekt 6 Monate nach Ausgabe, Einlösungsquote innerhalb von 6 Monaten nach Ausgabe

		Intensität			
		Niedrig		Hoch	
Stellenwert/ Informiertheit	Niedrig	Typ 1 (3.253)		Typ 3 (12.383)	
		Ausgabequote	3 %	Ausgabequote	3,95 %
		Treatment Effekt	6,61 %	Treatment Effekt	3,04 %
	Hoch	Typ 2 (5.812)		Typ 4 (9.548)	
		Ausgabequote	3,11 %	Ausgabequote	4,49 %
		Treatment - Effekt	7,20 %	Treatment - Effekt	5,11 %
		Einlösungsquote	14,08 %	Einlösungsquote	11,41 %
		Einlösungsquote	14,31 %	Einlösungsquote	11,41 %

In Westdeutschland (Tabelle 18) steigt der Treatment Effekt ebenfalls bei höherem Stellenwert und höherer Informiertheit. Gleichzeitig steigen die Einlösungsquoten. Da in Westdeutschland den Treatment Effekten ohnehin schon relativ geringe Einlösungsquoten gegenüber stehen, gibt es hier offenbar weniger Spielräume für eine Verbesserung. Eine Erhöhung der (Ausgabe)Intensität ist hier ebenfalls mit einer Erhöhung der Ausgabequote verbunden. Der Treatment Effekt sinkt hingegen weniger stark als im Osten und die Einlösungsquote steigt sogar geringfügig an.

Tabelle 18: Regionaler IAW-Typ in Westdeutschland: Ausgabequote, Treatment Effekt 6 Monate nach Ausgabe, Einlösungsquote innerhalb von 6 Monaten nach Ausgabe

		Intensität			
		Niedrig		Hoch	
Stellenwert/ Informiertheit	Niedrig	Typ 1 (13.713)		Typ 3 (7.481)	
		Ausgabequote	2,3%	Ausgabequote	2,58%
		Treatment Effekt	3,84%	Treatment Effekt	3,53%
	Hoch	Typ 2 (7.647)		Typ 4 (6.199)	
		Ausgabequote	1,63%	Ausgabequote	2,81%
		Treatment Effekt	6,25%	Treatment Effekt	4,86%
		Einlösungsquote	5,56%	Einlösungsquote	5,76%
		Einlösungsquote	8,30%	Einlösungsquote	8,60%

Inwiefern die Variable Stellenwert/Informiertheit ursächlich für die Unterschiede in den Treatment Effekten verantwortlich ist, soll später untersucht werden. Es ist nicht ausgeschlossen, dass die Unterschiede auch von der Zusammensetzung der Teilnehmergruppen in den Agenturtypen abhängen. Eine noch vorläufige Schätzung (nicht veröffentlicht) deutet darauf hin, dass die soziodemografische Struktur sehr wichtig ist.

Im Folgenden werden für jedes Arbeitsamt die drei Kennzahlen Ausgabequote, Einlösungsquote in den 6 Monaten nach Ausgabe und geschätzter Treatment Effekt 6 Monate nach Ausgabe dargestellt (Tabellen 19 und 20). Die geschätzten Effekte sind mit großer Vorsicht zu interpretieren. Dies gilt umso mehr, je kleiner die Zahl der Ausgaben ist, da dann der statistische Standardfehler relativ hoch ist.

Abgesehen von der statistischen Ungenauigkeit können die Unterschiede auf mehrere Faktoren zurückzuführen sein. Zunächst setzen sich die Bestände an anspruchsberechtigten Arbeitslosen in den einzelnen Agenturen unterschiedlich zusammen. In einer Agentur, in der relativ viele Arbeitslose gemeldet sind, für die der VGS einen höheren Effekt aufweist (z. B. jüngere Arbeitslose), sind die durchschnittlichen Treatment Effekte tendenziell höher. Weiterhin kann der durchschnittliche Effekt bei einer Agentur von der Arbeitsmarktlage im Bezirk abhängen. Schließlich hängt der durchschnittliche Effekt davon ab, ob die Agentur beim Einsatz des Instrumentes sowie die privaten Vermittler bei ihrer Tätigkeit eine gute Arbeit machen. Für die beiden zuerst genannten Einflüsse kann mit den verfügbaren Daten kontrolliert werden. Dadurch könnten die beiden zuletzt genannten Einflüsse gemeinsam isoliert werden. Auch wenn diese Schät-

zungen mit großen statistischen Unschärfen behaftet wären, könnten sie vermutlich wertvolle Hinweise für die Agenturmitarbeiter vor Ort liefern, wo noch Verbesserungspotenzial besteht. Dies wird in Zukunft genauer zu untersuchen sein.

Tabelle 19: Kennzahlen nach Agentur für Ostdeutschland, VGS-Ausgaben im Mai und Juni 2003, Einlösungen innerhalb von 6 Monaten und Treatment Effekt nach 6 Monaten

Agentur	Anzahl der aus- gegebenen VGS	Anteil der Einlösungen	Geschätzter Treatment Effekt
AA Neubrandenburg	1165	12,70 %	4,81 %
AA Rostock	1007	13,11 %	8,04 %
AA Schwerin	1040	11,44 %	3,56 %
AA Stralsund	830	11,33 %	1,20 %
AA Cottbus	2230	7,22 %	-0,72 %
AA Eberswalde	847	14,40 %	4,01 %
AA Frankfurt (Oder)	1042	10,65 %	3,26 %
AA Neuruppin	1214	11,86 %	9,14 %
AA Potsdam	1484	10,71 %	3,17 %
AA Dessau	1505	8,17 %	0,00 %
AA Halberstadt	215	17,21 %	6,05 %
AA Halle	1429	10,29 %	0,07 %
AA Magdeburg	992	14,82 %	12,40 %
AA Merseburg	556	16,01 %	8,09 %
AA Sangerhausen	473	15,01 %	9,73 %
AA Stendal	390	15,90 %	6,67 %
AA Wittenberg	240	13,33 %	9,58 %
AA Altenburg	146	17,81 %	4,79 %
AA Annaberg-Buchholz	522	20,50 %	11,88 %
AA Bautzen	1220	12,87 %	1,97 %
AA Chemnitz	798	13,78 %	10,65 %
AA Dresden	1287	8,24 %	4,97 %
AA Leipzig	2072	12,31 %	2,46 %
AA Oschatz	577	18,37 %	5,03 %
AA Pirna	406	10,59 %	4,19 %
AA Plauen	366	12,84 %	9,29 %
AA Riesa	376	15,69 %	11,44 %
AA Zwickau	764	16,62 %	9,55 %
AA Erfurt	895	14,30 %	7,60 %
AA Gera	585	11,97 %	3,93 %
AA Gotha	1149	10,18 %	-2,44 %
AA Jena	730	11,37 %	8,90 %
AA Nordhausen	611	15,55 %	12,27 %
AA Suhl	622	18,01 %	12,38 %

Tabelle 20: Kennzahlen nach Agentur für Westdeutschland, VGS-Ausgaben im Mai und Juni 2003, Einlösungen innerhalb von 6 Monaten und Treatment Effekt nach 6 Monaten

Agentur	Anzahl der aus- gegebenen VGS	Anteil der Einlösungen	Geschätzter Treatment Effekt
AA Bad Oldesloe	186	9,68 %	4,84 %
AA Elmshorn	210	4,76 %	5,24 %
AA Flensburg	323	7,43 %	2,79 %
AA Hamburg	1189	6,56 %	3,70 %
AA Heide	136	13,97 %	5,15 %
AA Kiel	92	8,70 %	4,35 %
AA Lübeck	162	12,35 %	1,85 %
AA Neumünster	145	11,03 %	12,41 %
AA Braunschweig	319	3,13 %	0,63 %
AA Bremen	243	7,00 %	10,70 %
AA Bremerhaven	118	9,32 %	9,32 %
AA Celle	113	12,39 %	7,96 %
AA Emden	355	3,66 %	-3,66 %
AA Goslar	126	6,35 %	2,38 %
AA Göttingen	308	1,62 %	5,84 %
AA Hameln	139	12,95 %	3,60 %
AA Hannover	1404	4,27 %	1,00 %
AA Helmstedt	163	11,04 %	4,29 %
AA Hildesheim	162	7,41 %	0,00 %
AA Leer	82	6,10 %	15,85 %
AA Lüneburg	364	6,87 %	-0,27 %
AA Nienburg	71	18,31 %	4,23 %
AA Nordhorn	83	2,41 %	6,02 %
AA Oldenburg	295	8,14 %	8,14 %
AA Osnabrück	170	13,53 %	9,41 %
AA Stade	85	14,12 %	8,24 %
AA Uelzen	43	9,30 %	6,98 %
AA Vechta	165	13,33 %	7,27 %
AA Verden	112	6,25 %	2,68 %
AA Wilhelmshaven	73	16,44 %	12,33 %
AA Aachen	109	11,93 %	15,60 %
AA Ahlen	207	4,83 %	12,08 %
AA Bergisch Gladbach	164	9,15 %	5,49 %
AA Bielefeld	3023	0,56 %	-1,16 %
AA Bochum	231	10,39 %	9,09 %
AA Bonn	324	9,26 %	4,32 %
AA Brühl	280	7,14 %	8,57 %
AA Coesfeld	93	16,13 %	15,05 %
AA Detmold	104	11,54 %	19,23 %
AA Dortmund	348	6,90 %	11,49 %
AA Düren	51	3,92 %	3,92 %
AA Düsseldorf	461	7,81 %	6,29 %
AA Duisburg	253	9,88 %	9,49 %
AA Essen	114	5,26 %	12,28 %
AA Gelsenkirchen	355	7,61 %	9,58 %
AA Hagen	222	6,31 %	12,16 %

Agentur	Anzahl der aus- gegebenen VGS	Anteil der Einlösungen	Geschätzter Treatment Effekt
AA Hamm	114	17,54 %	17,54 %
AA Herford	225	5,33 %	4,00 %
AA Iserlohn	73	1,37 %	5,48 %
AA Köln	252	9,52 %	9,13 %
AA Krefeld	186	4,30 %	6,45 %
AA Meschede	57	14,04 %	5,26 %
AA Mönchengladbach	343	6,71 %	12,83 %
AA Münster	50	8,00 %	8,00 %
AA Oberhausen	317	14,83 %	7,26 %
AA Paderborn	50	8,00 %	8,00 %
AA Recklinghausen	179	7,82 %	9,50 %
AA Rheine	134	7,46 %	8,96 %
AA Siegen	41	21,95 %	17,07 %
AA Soest	123	8,13 %	12,20 %
AA Solingen	76	2,63 %	7,89 %
AA Wesel	145	14,48 %	7,59 %
AA Wuppertal	343	2,33 %	5,83 %
AA Bad Hersfeld	72	9,72 %	19,44 %
AA Darmstadt	419	5,97 %	2,15 %
AA Frankfurt	606	4,29 %	1,65 %
AA Fulda	153	9,15 %	3,92 %
AA Giessen	270	5,56 %	4,81 %
AA Hanau	290	7,24 %	2,07 %
AA Kassel	748	3,48 %	0,40 %
AA Korbach	382	3,66 %	-3,93 %
AA Limburg	138	6,52 %	3,62 %
AA Marburg	158	4,43 %	1,90 %
AA Offenbach	92	11,96 %	3,26 %
AA Wetzlar	52	11,54 %	3,85 %
AA Wiesbaden	287	4,88 %	4,18 %
AA Bad Kreuznach	53	7,55 %	9,43 %
AA Kaiserslautern	148	14,19 %	7,43 %
AA Koblenz	234	1,28 %	0,43 %
AA Ludwigshafen	74	6,76 %	6,76 %
AA Mainz	168	6,55 %	9,52 %
AA Mayen	162	3,70 %	3,70 %
AA Montabaur	333	2,40 %	3,00 %
AA Neunkirchen	102	18,63 %	1,96 %
AA Landau	95	9,47 %	7,37 %
AA Neuwied	404	6,44 %	2,23 %
AA Pirmasens	236	20,34 %	-1,69 %
AA Saarbrücken	221	11,76 %	7,24 %
AA Saarlouis	95	8,42 %	17,89 %
AA Trier	82	8,54 %	13,41 %
AA Aalen	98	10,20 %	7,14 %
AA Balingen	30	6,67 %	-3,33 %
AA Freiburg	150	6,00 %	0,00 %
AA Göppingen	247	4,05 %	5,26 %
AA Heidelberg	69	10,14 %	5,80 %
AA Heilbronn	570	2,81 %	-0,53 %

Agentur	Anzahl der aus- gegebenen VGS	Anteil der Einlösungen	Geschätzter Treatment Effekt
AA Karlsruhe	91	7,69 %	14,29 %
AA Konstanz	107	7,48 %	5,61 %
AA Lörrach	80	12,50 %	-1,25 %
AA Ludwigsburg	70	10,00 %	11,43 %
AA Mannheim	82	4,88 %	2,44 %
AA Nagold	27	7,41 %	3,70 %
AA Offenburg	37	13,51 %	-5,41 %
AA Pforzheim	27	7,41 %	37,04 %
AA Rastatt	151	9,27 %	6,62 %
AA Ravensburg	73	5,48 %	-4,11 %
AA Reutlingen	101	16,83 %	4,95 %
AA Rottweil	22	9,09 %	0,00 %
AA Waiblingen	58	12,07 %	1,72 %
AA Schwäbisch Hall	33	6,06 %	6,06 %
AA Stuttgart	314	2,55 %	5,73 %
AA Tauberbischofsheim	18	5,56 %	16,67 %
AA Ulm	58	10,34 %	3,45 %
AA Villingen-Schwenningen	36	2,78 %	0,00 %
AA Ansbach	30	6,67 %	13,33 %
AA Aschaffenburg	38	5,26 %	18,42 %
AA Bamberg	149	2,01 %	-2,01 %
AA Bayreuth	55	12,73 %	25,45 %
AA Coburg	95	5,26 %	3,16 %
AA Hof	74	28,38 %	5,41 %
AA Nürnberg	267	5,24 %	5,99 %
AA Regensburg	70	12,86 %	5,71 %
AA Schwandorf	107	10,28 %	-0,93 %
AA Schweinfurt	444	4,96 %	0,45 %
AA Weiden	45	13,33 %	6,67 %
AA Weissenburg	123	4,07 %	-5,69 %
AA Würzburg	455	0,66 %	-0,88 %
AA Augsburg	104	15,38 %	17,31 %
AA Deggendorf	90	8,89 %	-3,33 %
AA Donauwörth	19	15,79 %	10,53 %
AA Freising	29	10,34 %	-10,34 %
AA Ingolstadt	171	5,85 %	8,77 %
AA Kempten	29	17,24 %	24,14 %
AA Landshut	81	7,41 %	1,23 %
AA Memmingen	49	4,08 %	-12,24 %
AA München	212	11,32 %	14,62 %
AA Passau	52	9,62 %	5,77 %
AA Pfarrkirchen	16	25,00 %	6,25 %
AA Rosenheim	69	18,84 %	-2,90 %
AA Traunstein	21	4,76 %	23,81 %
AA Weilheim	28	14,29 %	-3,57 %
AA Berlin Süd	1190	6,97 %	3,87 %
AA Berlin Südwest	496	6,25 %	2,22 %
AA Berlin Nord	945	8,36 %	9,63 %
AA Berlin Mitte	370	8,65 %	8,65 %
AA Berlin Ost	1571	10,18 %	6,37 %

Neben der Differenzierung nach einzelnen Merkmalen erfolgt auch eine Differenzierung nach dem geschätzten Propensity Score der Arbeitslosen. Dabei werden zunächst Quartile des Propensity Scores gebildet. Für die Teilnehmer innerhalb eines Quartils wird dann jeweils der durchschnittliche geschätzte Treatment Effekt sowie der Anteil der Einlösungen nach ein bis sechs Monaten berechnet. Die Ergebnisse sind in Tabellen 21 und 22 dargestellt.

Tabelle 21: ATT differenziert nach Propensity Score für Ostdeutschland

	<i>Verbleib in regulärer Beschäftigung</i>				<i>Für Teilnehmer: Anteil der Einlösungen</i>
1. Quartil (7.446)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
4	7,52 %	2,89 %	4,63 %	0,36 %	7,05 %
5	7,91 %	3,36 %	4,55 %	0,38 %	7,41 %
6	8,02 %	3,57 %	4,45 %	0,38 %	7,68 %
2. Quartil (7.446)					
4	13,50 %	8,90 %	4,59 %	0,52 %	10,23 %
5	14,21 %	9,60 %	4,61 %	0,54 %	10,68 %
6	14,17 %	9,66 %	4,51 %	0,54 %	11,11 %
3. Quartil (7.446)					
4	20,04 %	14,72 %	5,32 %	0,63 %	13,12 %
5	20,76 %	15,69 %	5,08 %	0,65 %	13,79 %
6	20,71 %	15,43 %	5,28 %	0,64 %	14,25 %
4. Quartil (7.447)					
4	25,66 %	20,49 %	5,17 %	0,72 %	14,62 %
5	26,63 %	21,42 %	5,21 %	0,73 %	15,54 %
6	25,78 %	20,83 %	4,96 %	0,72 %	15,93 %

Für Ostdeutschland ist zu erkennen, dass die Treatment Effekte bei den Arbeitslosen mit dem geringsten Propensity Score (1.Quartil) ebenfalls am geringsten sind. Bis zum dritten Quartil steigen die geschätzten Effekte, beim vierten Quartil sind die Effekte wieder etwas geringer. Gleichzeitig ist jedoch festzustellen, dass der Anteil der Einlösungen nach 6 Monaten noch wesentlich stärker mit dem Propensity Score ansteigt, von 7,7 Prozent im ersten auf 15,9 Prozent im vierten Quartil. Das zahlenmäßige Verhältnis vom Treatment Effekt zum Kosten treibenden Anteil der Einlösungen nach 6 Monaten ist im ersten Quartil somit am günstigsten und im vierten Quartil am ungünstigsten. Insgesamt deutet dies darauf hin, dass diejenigen Gruppen unter den Arbeitslosen häufiger einen VGS erhalten, bei denen

die Wahrscheinlichkeit der Einlösung am höchsten ist. Diese höhere Einlösungswahrscheinlichkeit scheint mit höheren positiven Treatment Effekten verbunden, aber in noch stärkerem Maße mit mehr Mitnahmen.

In Westdeutschland sind diese Zusammenhänge anders. Sowohl die geschätzten Treatment Effekte der Teilnehmer als auch der Anteil der Einlösungen nehmen bei höherem Propensity Score ab. Offenbar funktioniert der Selektionsmechanismus in Westdeutschland anders, was zu geringeren Mitnahmeeffekten führt.

Tabelle 22: ATT differenziert nach Propensity Score für Westdeutschland

	<i>Verbleib in regulärer Beschäftigung</i>				<i>Für Teilnehmer: Anteil der Einlösungen</i>
1. Quartil (8.150)					
Dauer nach Analysezeitpunkt (Monate)	Teilnehmer	Kontrollgruppe	Differenz	Std.fehler	
4	13,87 %	8,99 %	4,87 %	0,50 %	7,41 %
5	14,56 %	9,73 %	4,83 %	0,51 %	7,74 %
6	14,93 %	10,44 %	4,49 %	0,52 %	7,95 %
2. Quartil (8.150)					
4	16,90 %	12,53 %	4,37 %	0,56 %	6,72 %
5	18,04 %	13,72 %	4,32 %	0,58 %	6,96 %
6	18,40 %	14,18 %	4,22 %	0,58 %	7,17 %
3. Quartil (8.150)					
4	17,02 %	14,04 %	2,98 %	0,58 %	6,26 %
5	18,40 %	15,02 %	3,39 %	0,60 %	6,65 %
6	19,08 %	15,13 %	3,95 %	0,60 %	6,99 %
4. Quartil (8.150)					
4	16,63 %	14,43 %	2,20 %	0,60 %	4,40 %
5	17,69 %	15,87 %	1,83 %	0,62 %	4,69 %
6	18,21 %	16,00 %	2,21 %	0,62 %	4,87 %

4.5 Sensitivitätsanalysen

Es wurden u. a. die folgenden Sensitivitätsanalysen durchgeführt:

- Der Treatment Effekt wurde auch für diejenigen Arbeitslosen mit VGS-Ausgabe geschätzt, die laut BewA-Daten nicht anspruchsberechtigt gewesen wären. In diesem Fall müssen auch die nicht anspruchsberechtigten Nichtteilnehmer im Sample belassen werden. Um dies beim Matching zu berücksichtigen, wird zusätzlich das Merkmal „Anspruchsberechtigung“ als Kovariate verwendet. Die Zahl der Teilnehmer steigt bei dieser Modifikation auf 31.326 im Osten und 34.561 im Westen. Der

Treatment Effekt im Osten bleibt nahezu unverändert bei 4,7 Prozentpunkten, derjenige im Westen steigt hingegen signifikant auf 4,6 Prozentpunkte.

- Beim Matching wird die Kovariate „Dauer des letzten Beschäftigungsverhältnisses“ weggelassen. Diese Variable aus dem BewA fehlt bei einem Teil der Beobachtungen. Da die Erwerbshistorie der letzten fünf Jahre zusätzlich aus der IEB gezogen wird, scheint es vertretbar zu sein, diese Variable wegzulassen. Die Zahl der Teilnehmer steigt bei dieser Modifikation auf 30.903 im Osten und auf 34.851 im Westen. Die Treatment Effekte steigen auf 5,2 Prozentpunkte im Osten und 4,4 Prozentpunkte im Westen.

5 Zusammenfassung

Für die Vermittlungsgutscheine, die im Mai und Juni 2003 ausgegeben wurden, wird die Wahrscheinlichkeit der Einmündung in eine reguläre Beschäftigung innerhalb von sechs Monaten abgeschätzt. Der für die Gruppe der Arbeitslosen, die einen Vermittlungsgutschein erhalten haben, geschätzte durchschnittliche Maßnahmeeffekt beträgt den Ergebnissen zu Folge 4,8 Prozentpunkte in Ostdeutschland und 3,7 Prozentpunkte in Westdeutschland.

Die Wirkung der Vermittlungsgutscheine unterscheidet sich nach beobachteten Eigenschaften. Bei jüngeren Arbeitslosen konnten wir einen höheren Maßnahmeneffekt finden als bei älteren. Je höher die Vermittlungsprämie ist, desto geringer ist der Maßnahmeneffekt, was auf die damit verbundene höhere Arbeitslosigkeitsdauer zurückzuführen sein dürfte. Der höhere Anreiz für den privaten Vermittler zur Vermittlung durch die höhere Prämie kann dies offenbar nicht kompensieren. Ferner scheint das Instrument in Ostdeutschland bei Männern eine höhere Wirkung als bei Frauen zu entfalten. Eine geschlechtsspezifische unterschiedliche Wirkung konnte im Westen nicht festgestellt werden.

Nach IAB-Strategietypen getrennt wird der niedrigste Maßnahmeeffekt für den Strategietyp V (Westdeutschland mit guter Arbeitsmarktlage) gemessen. Dies ist insofern bemerkenswert, als dass in diesem Strategietyp die beobachteten Wiedereingliederungsquoten am höchsten sind. In diesem Sinne deutet die ökonometrische Analyse darauf hin, dass beobachtete Wiedereingliederungsquoten die Wirkung eines Instruments verzerrt wie-

dergeben können (in dem genannten Falle überschätzen). Aus der IAW-Typisierung scheint folgendes Ergebnis hervorhebenswert: Die Wirksamkeit ist in den Agenturen höher, in denen dem Instrument ein hoher Stellenwert beigemessen wird und in denen die Mitarbeiter gut über (mögliche) Missbrauchsfälle informiert sind. Eine höhere Intensität hinsichtlich des organisatorischen Einsatzes bei der Umsetzung des Instrumentes sowie hinsichtlich der Ausgabe von Vermittlungsgutscheinen scheint hingegen eher mit geringeren Effekten verbunden zu sein. Anscheinend ist die zunehmende Intensität mit einer abnehmenden zusätzlichen Wirkung verbunden.

Die Mikroanalyse legt die Schlussfolgerung nahe, dass die Vermittlungsgutscheine positive direkte Effekte entfalten, von denen einige Gruppen von arbeitslosen Teilnehmern stärker profitieren als andere. Es gibt insofern bedeutsame individuelle und regionale Maßnahmenheterogenität. Individuelle und regionale Eigenschaften, welche die Teilnahme befördern, begünstigen nicht zwangsläufig auch den Erfolg.

Um das Ausmaß der Selektionsverzerrung zu erkennen, sind die kausalen Effekte der Matching-Analyse mit den „naiven“ Ergebnissen des Projektmonitoring (Teil I des Forschungsberichtes) zu vergleichen. Die im Projektmonitoring gemessene Differenz im Vergleich zu allen Arbeitslosen beträgt 7-9 Prozentpunkte für die westdeutschen Strategietypen, der kausale Effekt beträgt jedoch nur 3,7 Prozentpunkte. Ein ähnliches Bild zeigt sich für Ostdeutschland: Im Projektmonitoring wurden 10 Prozentpunkte gemessen, der kausale Effekt beträgt dagegen nur 4,8 Prozentpunkte. Damit zeigt sich die Überschätzung der tatsächlichen Wirkung der Vermittlungsgutscheine durch die Positivauslese bei den Arbeitslosen, an die tatsächlich ein Vermittlungsgutschein ausgegeben wurde.

Literatur

Almus, M./Engeln, J./Lechner, M./Pfeiffer, F. und H. Spengler (1998): Die gemeinnützige Arbeitnehmerüberlassung in Rheinland-Pfalz - eine ökonomische Analyse des Wiedereingliederungserfolges, *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung (MittAB)* 31, 558-575.

Almus, M./Engeln, J./Lechner, M./Pfeiffer, F. und H. Spengler (1999): *Wirkungen gemeinnütziger Arbeitnehmerüberlassung in Rheinland-Pfalz*, BeitrAB 225, Nürnberg.

- Blundell, R. und M. Costa Dias (2002): Alternative approaches to evaluation in empirical microeconomics, *Portuguese Economic Journal*, Vol. 1 Issue 2, 91-116.
- Cochran, W. und D. Rubin (1973): Controlling Bias in Observational Studies: A Review, *Sankhya, Series A*, 35, 417-446.
- Dehejia, R.H. (2003): Practical Propensity Score Matching. A Reply to Smith and Todd, unpublished.
- Dehejia, R.H. und S. Wahba (2002): Propensity Score Matching Methods for Nonexperimental Causal Studies, *Review of Economics and Statistics* 84(1), 151-161.
- Dehejia, R.H. und S. Wahba (1999): Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs, *Journal of American Statistical Association*, 94, 1053-1062.
- Härdle, W. (1990): *Applied Nonparametric Regression*, *Econometric Society Monographs No. 19*, Cambridge (Cambridge University Press).
- Hagen, T und V. Steiner (2000): Von der Finanzierung der Arbeitslosigkeit zur Förderung von Arbeit, *ZEW-Wirtschaftsanalysen*, Bd.51, Baden-Baden.
- Heckman, J.J./Ichimura, H./Smith, J. und P. Todd (1996): Sources of selection bias in evaluating social programs: An interpretation of conventional measures and evidence on the effectiveness of matching as a program evaluation method, *Proceeding of the National Academy of Sciences* 93(23), 13416-13420.
- Heckman, J.J./Ichimura, H./Smith, J. und P. Todd (1998): Characterizing Selection Bias Using Experimental Data, *Econometrica* 66, 1017-1098.
- Heckman, J.J./Ichimura, H. und P. Todd (1997): Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, *Review of Economic Studies* 64, 605-654.
- Heckman, J.J./Ichimura, H. und P. Todd (1998): Matching as an Econometric Evaluation Estimator", *Review of Economic Studies*, 65, 261-294.
- Heckman, J.J./LaLonde, R.J. und J.A. Smith (1999): The Economics and Econometrics of Active Labour Market Programs, in: Ashenfelter, O. and D. Card (Hrsg.): *Handbook of Labor Economics*, Vol. 3a, 1865-2097.
- Heckman, J.J. und R. Robb (1985): Alternative Methods for Evaluating the Impact of Interventions, *Journal of Econometrics*, Vol. 30, Issue 1/2, 239-262.
- LaLonde, R.J. (1986): Evaluating the Econometric Evaluations of Training Programs, *American Economic Review*, 76, 604-620.

- Lechner, M. (1995): Effects of Continuous Off-the-Job Training in East Germany after Unification. ZEW Discussion Paper, Mannheim.
- Lechner, M. (1998): Training the East German Labour Force, Microeconomic Evaluations of Continuous Vocational Training after Unification, Heidelberg.
- Lechner, M. (1999): Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification. *Journal of Business & Economic Statistics*, Jan. 1999, Vol. 17, No. 1.
- Lechner, M. und F. Pfeiffer (2001): Econometric Evaluation of Labour Market Policies, ZEW Economic Studies 13, Heidelberg.
- Lechner, M. (2002): Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods. *Journal of the Royal Statistical Society*, 2002, 165, part 1.
- Rosenbaum, P.R. und D.B. Rubin (1983): The Central role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika* 70(1), 1-55.
- Rosenbaum, P.R. und D.B. Rubin (1985): Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score, *The American Statistician*, Feb. 1985, Vol. 39, No. 1.
- Rubin, D. (1980): Comment of Badu, D. – Randomization Analysis of Experimental Data: The Fisher Randomization Test, *Journal of American Statistical Association* 75, 591-593.
- Sianesi, B. (2001): An Evaluation of the Swedish System of Active Labour Market Programmes in the 1990, IFS Working Paper 03/01.
- Sianesi, B. (2003): An Evaluation of the Swedish System of Active Labour Market Programmes in the 1990, IFS Working Paper 02/01.
- Smith, J. und P. Todd (2003): Rejoinder, unpublished.
- Smith, J. und P. Todd (2004): Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?, *Journal of Econometrics*, forthcoming.
- Ullrich, B. (2004): Der Vermittlungsgutschein als Weg aus der Arbeitslosigkeit? Ein theoriegeleitetes Evaluationskonzept der arbeitsmarktpolitischen Maßnahme, Universität Lüneburg, Arbeitsbericht Nr. A325.

In dieser Reihe sind zuletzt erschienen

Nr.	Autor(en)	Titel	Datum
1/2004	Sabine Hagemann, Werner Sörgel, Eberhard Wiedemann	Vermittlungsgutscheine nach § 421g SGB III - Zwischenergebnisse aus der Begleitforschung zur Vermittlung	9/2004
2/2004	Lutz Bellmann, Vera Dahms, Jürgen Wahse	IAB-Betriebspanel Ost - Ergebnisse der achten Welle 2003 – Teil I: Entwicklung und Struktur der Betriebe und Beschäftigten, Auszubildende	9/2004
3/2004	Lutz Bellmann, Vera Dahms, Jürgen Wahse	IAB-Betriebspanel Ost - Ergebnisse der achten Welle 2003 – Teil II: Personalpolitik, Betriebliche Flexibilität, Weiterbildung	9/2004
4/2004	Lutz Bellmann, Vera Dahms, Jürgen Wahse	IAB-Betriebspanel Ost - Ergebnisse der achten Welle 2003 – Teil III: Wirtschaftliche Lage der Betriebe, Öffentliche Förderung	9/2004
5/2004	Eugen Spitznagel, Susanne Wanger	Mehr Beschäftigung durch längere Arbeitszeiten? Ein Beitrag zu der Diskussion um eine generelle Erhöhung der Arbeitszeit	10/2004
6/2004	IAB-Autoren- gemeinschaft	Forschung zum SGB II des IAB: Die neuen Forschungsaufgaben im Überblick	12/2004
1/2005	Anja Heinze, Friedhelm Pfeiffer, Alexander Sperrmann, Henrik Winterhager, Amelie Wuppermann	Vermittlungsgutscheine - Zwischenergebnisse der Begleitforschung 2004 Teil I: Datenstruktur und deskriptive Analysen	3/2005
2/2005	Sabine Dann, Günther Klee, Martin Rosemann	Vermittlungsgutscheine - Zwischenergebnisse der Begleitforschung 2004 Teil II: Typisierung der Arbeitsagenturen	2/2005

Impressum

IAB *Forschungsbericht*
Nr. 3 / 2005

Herausgeber

Institut für Arbeitsmarkt- und Berufsforschung
der Bundesagentur für Arbeit
Weddigenstr. 20-22
D-90478 Nürnberg

Redaktion

Regina Stoll, Jutta Palm-Nowak

Technische Herstellung

Jutta Sebald

Rechte

Nachdruck – auch auszugsweise – nur mit
Genehmigung des IAB gestattet

Bezugsmöglichkeit

Volltext-Download dieses Forschungsberichtes
unter:

<http://doku.iab.de/forschungsbericht/2005/fb0305.pdf>

IAB im Internet

<http://www.iab.de>

Rückfragen zum Inhalt an

Henrik Winterhager, Tel. 0621/1235-280,
oder e-Mail: winterhager@zew.de