



Universität Potsdam

Julia Reilich

Bildungsrenditen in Deutschland

Eine nationale und regionale Analyse

Potsdamer Schriften zur Raumwirtschaft | 5
Prof. Dr. Klaus Schöler (Hrsg.)

Potsdamer Schriften zur Raumwirtschaft
Prof. Dr. Klaus Schöler (Hrsg.)

Potsdamer Schriften zur Raumwirtschaft | 5
Prof. Dr. Klaus Schöler (Hrsg.)

Julia Reilich

Bildungsrenditen in Deutschland

Eine nationale und regionale Analyse

Universitätsverlag Potsdam

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.de/> abrufbar.

Universitätsverlag Potsdam 2013

<http://verlag.ub.uni-potsdam.de/>

Am Neuen Palais 10, 14469 Potsdam

Tel.: +49 (0)331 977 2533 / Fax: 2292

E-Mail: verlag@uni-potsdam.de

Die Schriftenreihe **Potsdamer Schriften zur Raumwirtschaft** wird herausgegeben von Prof. Dr. Klaus Schöler.

ISSN (print) 2190-8702

ISSN (online) 2190-8710

Das Manuskript ist urheberrechtlich geschützt.

Online veröffentlicht auf dem Publikationsserver der
Universität Potsdam:

URL <http://pub.ub.uni-potsdam.de/volltexte/2013/6265/>

URN <urn:nbn:de:kobv:517-opus-62658>

<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-62658>

Zugleich gedruckt erschienen im Universitätsverlag Potsdam:

ISBN 978-3-86956-219-3

Geleitwort

Der fünfte Band der „Potsdamer Schriften zur Raumwirtschaft“, der Beitrag von Frau Dr. Julia Reilich mit dem Titel „Bildungsrenditen in Deutschland – Eine nationale und regionale Analyse“, setzt neue Akzente innerhalb der Schriftenreihe und erweitert das Spektrum in zweifacher Hinsicht. Waren die ersten drei Bände der Neuen Ökonomischen Geographie verpflichtet, so liegt nunmehr eine Arbeit aus dem Forschungsgebiet der Arbeitsmarkt- und Bildungsökonomik vor; waren die bisherigen Beiträge überwiegend theoretisch-analytisch orientiert, so sehen wir nunmehr einen Forschungsbeitrag, der empirisch ausgerichtet ist. Somit stellt der fünfte Band eine willkommene methodische und thematische Ausweitung dar, ohne dabei das zentrale Anliegen, die Raumwirtschaft, zu vernachlässigen.

Das Buch beschäftigt sich mit einem theoretisch und wirtschaftspolitisch überaus wichtigen Thema, der Frage, wie hoch ist die Verzinsung des individuell gebildeten Humankapitals? Diese Frage ist nicht neu, jedoch werden in dieser Arbeit – neben anderen – zwei neue Antworten gegeben, die einen wichtigen Forschungsfortschritt auf diesem Gebiet darstellen.

1. Es wird gezeigt, daß die Höhe der Bildungsrendite bei gleichem Datenmaterial von den angewendeten statistischen Verfahren abhängt und man nicht von der Bildungsrendite sprechen kann, sondern von einem Werte-Korridor.
2. Erstmals werden regional unterschiedlichen Bildungsrenditen für Deutschland auf der Basis der Bundesländer ermittelt, wobei Unterschiede durch historische Einflüsse (alte Bundesländer vs. Beitrittsgebiete) deutlich werden.

Es ist ein unschätzbare Vorteil dieses Buches, die statistischen Tests in einer einfachen und klaren Sprache darzustellen, die komplexen ökonometrischen Verfahren dem Leser nahe zu bringen und die Ergebnisse in allgemeinverständlichen Zusammenfassungen zu präsentieren. Auch wird die Orientierung innerhalb der Arbeit durch eine klare, nachvollziehbare Gliederung erleichtert. Da die Forschungsergebnisse, die in diesem Buch veröffentlicht werden, einen wichtigen Beitrag zur Bildungsökonomik und zur empirisch orientierten

tierten Raumwirtschaftstheorie leisten und eine wertvolle Informationsbasis für politische Entscheidungsträger in Bund, Länder und Gemeinden bilden, wünsche ich dem Werk eine weite Verbreitung.

Potsdam, im November 2012

Klaus Schöler

Vorwort

In vielen Studien wird die Bundesrepublik Deutschland als homogener Standort untersucht. So bisher auch bezüglich der Frage von der Verzinsung von allgemeinem Humankapital. Es ergeben sich jedoch diverse Unterscheidungsmerkmale zum Beispiel hinsichtlich der Wirtschaftskraft, der Geschichte, der geografischen Gegebenheiten. Warum sollte es daher in Deutschland eine einheitliche Wertschätzung von Bildung geben? Der Vergleich internationaler Studien zeigt, dass zwischen ähnlich entwickelten Ländern homogene Bildungsrenditen geschätzt werden. Basierend auf der Überlegung regionaler Disparitäten war es mein Ziel, Deutschlands Regionen, hinsichtlich dieser durchschnittlichen Entlohnung für Wissen zu untersuchen und die Ergebnisse vergleichbar zu machen. Diese Arbeit verfasste ich während meiner Zeit als wissenschaftliche Mitarbeiterin an der Universität Potsdam am Lehrstuhl für Volkswirtschaftslehre, insbesondere Wirtschaftstheorie.

Es ist mir an dieser Stelle eine angenehme Pflicht jenen zu danken, die zum Entstehen dieser Arbeit maßgeblich beigetragen haben. Herrn Prof. Dr. Klaus Schöler, meinem Doktorvater, danke ich für seine stetige Unterstützung in den letzten Jahren, sei es in fachlicher oder moralischer Hinsicht oder sei es als stets sehr korrekter Vorgesetzter. Auch meinem Zweitgutachter Herrn Prof. Dr. Hans-Gerhard Strohe gilt mein Dank. Er hat mir den ersten Einblick in wissenschaftliches Arbeiten ermöglicht und mich durch viele Diskussion über die sinnvolle Verwendung von Daten sensibilisiert. Den Vorsitz der Prüfungskommission übernahm Herr Prof. Dr. Dieter Holtmann. Ihm danke ich hierfür und für diverse konstruktive Zusammenkünfte. Herrn Prof. Dr. Norbert Eickhof danke ich für vielfältige Diskussionen sowie die Bereitschaft die Prüfungskommission zu bereichern.

Insbesondere gilt mein Dank Herrn Prof. Dr. Gert Kneis, ohne den ich wohl nie auf die Idee gekommen wäre, das Projekt Doktorarbeit anzustreben. Herrn Prof. Scott Drewianka, Phd danke ich für die Inspiration und stetige Ansprechbereitschaft, trotz großer räumlicher Distanz. Ebenso danke ich Herrn Prof. Dr. Uwe Blien und Dr. Wolfgang Dauth für ihren Rat und ihre konstruktiver Kritik im Entstehen dieser Arbeit. Meinem Kollegen, Herrn Kai

Andree, gilt ein besonderer Dank. Er stand mir fachlich, wie persönlich Tag für Tag zur Seite. Herr Richard Landgraf half mir insbesondere beim Layout meiner Arbeit.

Weiterhin danke ich meinem Bruder Michael, meiner Schwägerin Kathlen, meinem Vater Stephan, meiner Mutter Christa Reilich sowie Frau Kristin Valerius und Frau Katja Cwejn, die durch ihre kritische Durchsicht des Manuskriptes stark zum Werden dieser Arbeit beigetragen haben. Abschließend danke ich meiner Familie, die mich mein Leben lang unterstützt hat und die viele steinige Wege geebnet hat.

Potsdam, im November 2012

Julia Reilich

Inhaltsverzeichnis

1	Einleitung	1
1.1	Fragestellung	1
1.2	Aufbau der Arbeit	3
2	Theoretische Fundierung	5
2.1	Der Einfluss von allgemeinem Humankapital – Das Grundmodell nach Jacob Mincer	7
2.1.1	Grundmodell	7
2.1.2	Grundmodell mit konstanter Zahl an Berufsjahren	10
2.1.3	Kritik	14
2.2	Einfluss von allgemeinem und firmenspezifischem Humankapital	16
2.2.1	Erweitertes Modell von Jacob Mincer	16
2.2.2	Kritik	21
2.3	Zusammenfassung	22
3	Stand der Forschung	23
4	Daten und Datenbeschreibung	29
4.1	Datensatzeingrenzungen	29
4.1.1	Allgemeiner Datensatz	29
4.1.2	Datensatz der Erwerbstätigen	30
4.1.3	Repräsentativitätsprüfung des Datensatzes	31
4.2	Beschreibung der Variablen und deskriptive Analyse	39
4.2.1	Variablen zur Schätzung der Lohngleichung	39
4.2.2	Variablen zur Schätzung der Bildungsdauer	49
4.2.3	Separierende Variablen des Datensatzes	53
4.2.4	Zusammenfassung	56

4.3	Korrelationsbeziehungen	56
4.3.1	Korrelationsbeziehungen zum Bruttostundenlohn	57
4.3.2	Korrelationsbeziehungen zur Bildungsdauer	58
4.3.3	Zusammenfassung	59
5	Schätzmethoden und Ermittlung der nationalen Bildungsrendite	61
5.1	Schätzverfahren – Methode der kleinsten Quadrate (OLS)	61
5.1.1	Modellbeschreibungen	69
5.1.2	Bildungsrenditen in Deutschland	73
5.1.3	Zusammenfassung	76
5.2	Mehrgleichungsmodelle zur Schätzung von Bildungsrenditen	77
5.2.1	Two-Stage Least Squares	78
5.2.2	Three-Stage Least Squares	95
5.2.3	Zusammenfassung	98
5.3	Bildungsrenditen nach verschiedenen Merkmalen	99
5.3.1	Kohorteneffekte	100
5.3.2	Öffentlicher Dienst versus Privatwirtschaft	102
5.3.3	Vollzeit- versus Teilzeitbeschäftigte	105
5.3.4	Zusammenfassung	107
5.4	Berücksichtigung der Erwerbslosen	108
5.4.1	Theoretische Fundierung	108
5.4.2	Bildungsrenditen für Erwerbstätige und ALG-I-Empfänger	110
5.4.3	Bildungsrenditen für alle Erwerbspersonen	119
5.4.4	Kritik	131
5.4.5	Zusammenfassung	133
6	Regionale Bildungsrenditen	135
6.1	Bildungsrenditen in Ost- und Westdeutschland	136
6.1.1	Deskriptive Statistiken	136
6.1.2	Ergebnisdarstellung	138
6.1.3	Zusammenfassung	141
6.2	Regionale Bildungsrenditen	141
6.2.1	Regionale Bildungsrenditen auf Basis separierter Stichproben	141
6.2.2	Regionale Bildungsrenditen auf Basis einer gesamtdeutschen Stichprobe	152
6.2.3	Vergleich der regionalen Bildungsrenditen	161

6.2.4	Kritik	163
6.2.5	Zusammenfassung	164
6.3	Regionale Bildungsrenditen unter Berücksichtigung der Erwerbslosen . .	165
6.3.1	Regionale Bildungsrenditen auf Basis separierter Stichproben . . .	165
6.3.2	Regionale Bildungsrenditen auf Basis einer gesamtdeutschen Stich- probe	170
6.3.3	Zusammenfassung	176
6.4	Vergleich der Schätzergebnisse und Zusammenfassung	176
6.4.1	Vergleich der verwendeten Methoden	177
6.4.2	Vergleich der geschätzten Bildungsrenditen	178
6.4.3	Kritik	181
6.5	Regionale Unterschiede in Deutschland	182
6.5.1	Bedeutung der regionalen Bildungsrendite	183
6.5.2	Kritik	189
6.5.3	Zusammenfassung	190
7	Schlussbetrachtung	193
	Literaturverzeichnis	199
	Anhang	211
A.1	Theoretische Herleitung der Mincer-Gleichung	211
A.2	Daten und Datenbeschreibung	212
A.2.1	Repräsentativitätsprüfung	213
A.2.2	Beschreibung der Variablen und Darstellung der Testergebnisse .	215
A.3	Schätzverfahren und Ermittlung der nationalen Bildungsrendite	220
A.3.1	Allgemeine Darstellung der OLS-Annahmen	220
A.3.2	Varianzinflationsfaktoren aus der OLS-Regression	226
A.3.3	Allgemeine Darstellung der IV-Annahmen	226
A.3.4	Multikollinearitätsprüfung für die Reduced-Form-Regression . . .	226
A.3.5	Ergebnisse der Reduced-Form-Regression für verschiedene Ko- horten	229
A.3.6	Berücksichtigung der Erwerbslosen	230
A.4	Regionale Analyse	231
A.4.1	Ergebnisse der Reduced-Form-Regression pro Bundesland	232

A.4.2	Varianzinflationsfaktoren aus der regionalen OLS-Regression mit Interaktionstermen	232
A.4.3	Ergebnisse der separierten OLS-Regression	235
A.4.4	Berücksichtigung der Erwerbslosen	240
A.4.5	Regionale Unterschiede – Zusammenhang zur Bildungsrendite . .	243

Abbildungsverzeichnis

2.1	Lebenszyklus bei Unterscheidung in generelles und firmenspezifisches Humankapital	5
2.2	Einkommensstrom für hohe und niedrige Qualifizierung	6
2.3	Lebenszyklus	8
2.4	Lebenszyklus für eine konstante Zahl an Berufsjahren	11
2.5	Zusammenhang zwischen der Bildungsrendite und der Zeitpräferenzrate und die Auswirkung auf die Bildungsentscheidung	13
2.6	Indexierung im Modell nach Mincer	18
4.1	Verteilung der Schulabschlüsse in Deutschland im Vergleich	34
4.2	Anteil männlicher Erwerbstätiger nach Altersgruppen im Vergleich	35
4.3	Häufigkeitsverteilungen des Familienstands der 30–60-Jährigen im Vergleich.	36
4.4	Verteilung der Erwerbstätigen auf die Bundesländer im Vergleich	37
4.5	Dauer der Ausbildung versus mittlerer Bruttostundenlohn	42
4.6	Mittlere Bildungsdauer pro Anzahl an Geschwistern	51
5.1	Vergleich der Dauer der Ausbildung mit dem mittleren logarithmierten Bruttostundenlohn	63
5.2	Optische Prüfung räumlicher Autokorrelation	66
5.3	Der Einfluss der Berücksichtigung der Variable „Fähigkeiten“	80
5.4	IV-Verfahren	81
5.5	Kontemporäre Korrelation bei Mehrgleichungsmodellen	96
5.6	Überblick zur Schätzung der Bildungsrenditen für Erwerbstätige und ALG-I-Empfänger	114
5.7	Relative Häufigkeitsverteilung der Schulabschlüsse der Erwerbslosen in Deutschland im Vergleich	125

5.8	Relative Häufigkeitsverteilung der Schulabschlüsse der Erwerbspersonen in Deutschland im Vergleich	126
6.1	Übersicht Kapitel 6	135
6.2	Vergleich relative Häufigkeiten auf Bundesländerebene	144
6.3	Mittelwertvergleich pro Bundesland	145
6.4	Bildungsrenditen pro Bundesland (3SLS-Methode)	147
6.5	95 %-Konfidenzintervalle für die Bildungsrenditen pro Bundesland (3SLS-Verfahren)	151
6.6	Bildungsrenditen pro Bundesland (IAT)	154
6.7	95 %-Konfidenzintervalle für die Bildungsrenditen pro Bundesland (IAT)	158
6.8	Regionale Bildungsrenditen im Vergleich	163
6.9	Grafische Ergebnisdarstellung der nach Bundesländern separierten Heckman-Regression	166
6.10	95 %-Konfidenzintervalle der Bildungsrenditen mit der nach Bundesländern separierten Heckman-Methode	170
6.11	Grafische Ergebnisdarstellung der Bildungsrenditen aus der Heckman-Regression mit IAT	172
6.12	95 %-Konfidenzintervalle für die mit dem Heckman-Verfahren ermittelten Bildungsrenditen pro Bundesland (IAT)	174
6.13	Bildungsrenditen im Vergleich	180
6.14	Regionales verfügbares Einkommen der Haushalte pro Einwohner und Bildungsrenditen	183
6.15	Regionale relative Standardabweichung des Bruttostundenlohns und Bildungsrenditen	185
6.16	Wanderungssalden der Bundesländer zwischen 2000 und 2006	186
6.17	Regionale Arbeitslosenquoten und Bildungsrenditen	187
6.18	Regionale Baulandpreise und Bildungsrenditen	188
A.1	Verteilung der Schulabschlüsse pro Bundesland, Teil 1	213
A.2	Verteilung der Schulabschlüsse pro Bundesland, Teil 2	214
A.3	Verteilung der Schulabschlüsse pro Bundesland, Teil 3	215
A.4	95 %-Konfidenzintervalle der separierten OLS-Regression auf Bundesländerebene im Zahlenstrahl	239

Tabellenverzeichnis

3.1	Bildungsrenditen in Deutschland Teil 1	24
3.2	Bildungsrenditen in Deutschland Teil 2	25
4.1	Reduktion aus der Grunddatei	30
4.2	Anzahl der gesamt generierten Jahre mit zugehörigem Abschluss	41
4.3	P-Werte aus den zweiseitigen Signifikanztests auf Gleichheit zwischen den mittleren Bruttostundenlöhnen pro Bildungsdauer	45
4.4	Deskriptive Statistiken pro Kohorte	53
4.5	Vergleich der Bruttostundenlöhne im privaten und öffentlichen Sektor für ausgewählte Bildungsdauern	55
4.6	Korrelationen zum logarithmierten Stundenlohn	57
4.7	Korrelationen zur Bildungsdauer	59
5.1	Variablenbezeichnung – Grundmodell.	70
5.2	Variablenbezeichnung – Persönliche Gegebenheiten.	71
5.3	Variablenbezeichnung – Unternehmensgröße.	72
5.4	Variablenbezeichnung – Wohnort.	72
5.5	Ergebnistabelle der OLS-Regression	74
5.6	IV-Verfahren – Modell 1.	82
5.7	IV-Verfahren – Modell 2.	83
5.8	IV-Verfahren – Modell 3.	84
5.9	IV-Verfahren – Modell 4.	84
5.10	IV-Verfahren – Modell 5.	85
5.11	IV-Verfahren – Modell 6.	86
5.12	Korrelationen zwischen dem Residuum der Lohn- und Bildungsgleichung pro Modellspezifikation	89
5.13	Ergebnistabelle der Reduced-Form-Regression Teil 1	90
5.14	Ergebnistabelle der Reduced-Form-Regression	91

5.15	Ergebnisse für die Schätzung der Lohngleichung mit der IV-Methode . . .	92
5.16	Ergebnistabelle der 3SLS-Regression	98
5.17	Ergebnistabelle der 3SLS-Regression für die Kohorten	101
5.18	95 %- Konfidenzintervalle für die Bildungsrenditen pro Kohorte	102
5.19	Ergebnistabelle der 3SLS-Regression für Arbeitnehmer im öffentlichen Dienst und in der Privatwirtschaft	103
5.20	95 %-Konfidenzintervalle für die Bildungsrendite im öffentlichen und pri- vaten Sektor	104
5.21	95 %-Konfidenzintervalle für die Bildungsrendite voll- und teilzeitbeschäf- tigter Arbeitnehmer	105
5.22	Ergebnistabelle der 3SLS-Regression für Voll- und Teilzeitbeschäftigte . .	106
5.23	Deskriptive Statistiken des erweiterten Datensatzes mit ALG-I-Empfängern gesamt und separat	113
5.24	Mittelwertvergleich des erweiterten Datensatzes mit ALG-I-Empfängern gesamt und separat	113
5.25	Variablenbezeichnung – Erweiterung um ALG-I-Empfänger.	116
5.26	Ergebnistabelle der Lohngleichung der 3SLS-Regression für den erwei- terten (Erwerbstätig und ALG I) Datensatz zusammen und getrennt sowie den ursprünglichen Datensatz im Vergleich	117
5.27	95 %-Konfidenzintervalle im Vergleich: separierter Datensatz	118
5.28	Deskriptive Statistiken des um die Erwerbslosen erweiterten Datensatzes gesamt und separat	123
5.29	P-Werte aus dem Vergleich der Bildungsdauern der Erwerbstätigen und Erwerbslosen	123
5.30	P-Werte aus dem Vergleich der Dauer der bisherigen Arbeitslosigkeit der Erwerbstätigen und Erwerbslosen	127
5.31	P-Werte aus dem Vergleich der Anzahl der Kinder im Haushalt der Er- werbstätigen und Erwerbslosen	128
5.32	Variablenbezeichnung mit vermutetem Zusammenhang zur Arbeitsmarkt- teilnahme.	129
5.33	Ergebnistabelle der Heckman-Regression	131
6.1	Regionaler Vergleich anhand relativer Häufigkeiten	137
6.2	Regionaler Vergleich der mittleren Bildungsdauer und des mittleren Brut- tostundenlohns.	138

6.3	Ergebnistabelle der 3SLS-Regression für Ost- und Westdeutschland . . .	139
6.4	95 %-Konfidenzintervalle für die Bildungsrendite in Ost- und Westdeutschland	140
6.6	Regionaler Vergleich anhand relativer Häufigkeiten pro Bundesland . . .	142
6.8	Mittelwertvergleich auf Bundesländerebene	146
6.9	Ergebnistabelle der 3SLS-Regression auf Bundesländerebene, Teil 1 . . .	148
6.10	Ergebnistabelle der 3SLS-Regression auf Bundesländerebene, Teil 2 . . .	149
6.12	95 %-Konfidenzintervalle aus der 3SLS-Regression auf Bundesländerebene	150
6.13	Variablenbezeichnung – Interaktionsterme.	153
6.14	Ergebnistabelle der OLS-Regression Interaktionstermen	156
6.16	95 %-Konfidenzintervalle aus der Regression mit Interaktionstermen für die Bundesländer	157
6.17	Ergebnisse des Wald-Tests auf Unterschiedlichkeit der Bildungsrenditen pro Bundesland	160
6.18	Regionale Bildungsrenditen im Vergleich	162
6.19	Ergebnistabelle der Heckman-Regression auf Bundesländerebene, Teil 1 .	167
6.20	Ergebnistabelle der Heckman-Regression auf Bundesländerebene, Teil 2 .	168
6.22	95 %-Konfidenzintervalle aus der Heckman-Regression separiert nach Bundesländern	169
6.23	Ergebnistabelle der Heckman-Regression mit IAT pro Bundesland	171
6.25	95 %-Konfidenzintervalle aus der Heckman-Regression mit Interaktionstermen für die Bundesländer	173
6.26	Ergebnisse des Wald-Tests auf Unterschiedlichkeit der Bildungsrenditen aus der Heckman-Regression (IAT)	175
6.27	Regionale Bildungsrenditen im Vergleich	179
A.1	P-Werte aus dem Levene-Test auf Homogenität der Varianzen pro Bildungsdauer	216
A.2	P-Werte aus dem Vergleich des mittleren Bruttostundenlohns separiert nach einzelnen erklärenden Variablen	217
A.3	P-Werte aus den Vergleichen der mittleren Bildungsdauern separiert nach einzelnen Instrumenten	218
A.4	P-Werte aus den Vergleichen des mittleren Bruttostundenlohn und der mittleren Bildungsdauern separiert nach Gruppen	219

A.5	P-Werte zum Vergleich mittlerer Bruttostundenlöhne ausgewählter Bildungsdauern im privaten und öffentlichen Sektor	220
A.6	Darstellung der OLS-Annahmen, Teil 1.	221
A.7	Darstellung der OLS-Annahmen, Teil 2.	222
A.8	Darstellung der OLS-Annahmen, Teil 3.	223
A.9	Darstellung der OLS-Annahmen, Teil 4.	224
A.10	Darstellung der OLS-Annahmen, Teil 5.	225
A.11	Varianzinflationsfaktoren aus der OLS-Regression	226
A.12	Darstellung der IV-Annahmen, Teil 1.	227
A.13	Darstellung der IV-Annahmen, Teil 2.	228
A.14	Varianzinflationsfaktoren der Reduced-Form-Regression aus der IV-Regression	229
A.15	Ergebnistabelle der Reduced-Form-Regression separiert nach Kohorten	230
A.16	Ergebnistabelle der Reduced-Form-Regression für den erweiterten (Erwerbstätig und ALG I) Datensatz zusammen und getrennt sowie den ursprünglichen Datensatz im Vergleich	231
A.19	Varianzinflationsfaktoren aus der gesamtdeutschen OLS-Regression	232
A.17	Ergebnistabelle der Reduced-Form-Regression auf Bundesländerebene Teil 1	233
A.18	Ergebnistabelle der Reduced-Form-Regression auf Bundesländerebene Teil 2	234
A.20	Ergebnistabelle der OLS-Regression auf Bundesländerebene, Teil 1	236
A.21	Ergebnistabelle der OLS-Regression auf Bundesländerebene, Teil 2	237
A.23	95 %-Konfidenzintervalle aus der separierten OLS-Regression auf Bundesländerebene	238
A.24	VIF für die separierte OLS-Regression auf Bundesländerebene	240
A.25	Ergebnistabelle der Probit-Regression separiert pro Bundesland, Teil 1	241
A.26	Ergebnistabelle der Probit-Regression separiert pro Bundesland, Teil 2	242
A.27	Ergebnisse der Probit-Regression der Heckman-Methode zur regionalen Analyse mit Interaktionstermen	243
A.28	Korrelationskoeffizienten zur regionalen Bildungsrendite	244

Abkürzungsverzeichnis

<i>2SLS</i>	Two-Stage Least Squares, Instrumentvariablenmethode
<i>3SLS</i>	Three-Stage Least Squares
<i>BLUE</i>	Best Linear Unbiased Estimator
<i>BP</i>	Breusch-Pagan
<i>DIW</i>	Deutsches Institut für Wirtschaftsforschung
<i>DWH</i>	Durbin-Wu-Hausmann
<i>GLS</i>	Generalized Least Squares
<i>IAB</i>	Institut für Arbeitsmarkt- und Berufsforschung
<i>IAT</i>	Interaktionsterm
<i>IV</i>	Instrumentvariablen
<i>IW</i>	Institut der Deutschen Wirtschaft Köln
<i>JB</i>	Jarque-Bera
<i>LM</i>	Lagrange-Multiplikator
<i>OLS</i>	Methode der kleinsten Quadrate
<i>P – Wert</i>	Wahrscheinlichkeitswert
<i>PSID</i>	Panel Study of Income Dynamics
<i>SOEP</i>	Sozio-Ökonomisches Panel
<i>SGB</i>	Sozialgesetzbuch
<i>VIF</i>	Varianzinflationsfaktor

Bundesländerabkürzungen ¹

<i>BW</i>	Baden-Württemberg
<i>BY</i>	Bayern
<i>BE</i>	Berlin
<i>BB</i>	Brandenburg
<i>HB</i>	Bremen
<i>HH</i>	Hamburg
<i>HE</i>	Hessen
<i>MV</i>	Mecklenburg-Vorpommern
<i>NI</i>	Niedersachsen
<i>NW</i>	Nordrhein-Westfalen
<i>RP</i>	Rheinland-Pfalz
<i>SL</i>	Saarland
<i>SN</i>	Sachsen
<i>ST</i>	Sachsen-Anhalt
<i>SH</i>	Schleswig-Holstein
<i>TH</i>	Thüringen

¹Nach ISO-3166-2-Code.

Symbolverzeichnis

a_i	Nichtarbeitseinkommen von Individuum i
ALD_i	Dauer der Arbeitslosigkeit von Individuum i
C_t	Opportunitätskosten aus entgangenem Einkommen zum Zeitpunkt t
d	Um d Jahre verringerte generelle Bildungsdauer
D_{AL_i}	Individuum i ist arbeitslos
$D_{BL_{il}}$	Individuum i wohnt in Bundesland l
D_{DDR_i}	Individuum i hat einen Abschluss aus der ehemaligen DDR
$D_{gro\beta_i}$	Individuum i arbeitet in einem Unternehmen mit mehr als 200 Mitarbeitern
D_{HB_i}	Vater von Individuum i hat eine hohe Bildung
$D_{l\ddot{a}ndl_i}$	Individuum i wohnt im ländlichen Gebiet
D_{MB_i}	Vater von Individuum i hat eine mittlere Bildung
$D_{Nicht-deutsch_i}$	Individuum i hat keine deutsche Staatsbürgerschaft
$D_{Pendler_i}$	Individuum i pendelt mindestens einmal die Woche
$D_{T\ddot{a}tig\ddot{e}rl_i}$	Ausgeübte berufliche Tätigkeit von Individuum i wurde zuvor erlernt
ε	Fehlerterm der Lohngleichung
E_t	Potentielles Einkommen in Periode t
Erf_i	Potentielle Berufserfahrung von Individuum i
GE	Grenzeinkommen
GK	Grenzkosten
GN	Grenznutzen
h_i	Wöchentliche Arbeitsstunden von Individuum i
HC	Humankapital
IAT	Interaktionsterm

Symbolverzeichnis

k_t	Investitionsquote zum Zeitpunkt t
$\hat{\lambda}_i$	Korrekturterm für Individuum i für das Heckman-Verfahren
μ	Mittelwert eines Merkmals in der Grundgesamtheit
η	Fehlerterm aus der Reduced-Form-Regression
$N_{Buecher}$	Anzahl der Bücher im Haushalt
N_{Brued}	Anzahl der Brüder
N_{Schw}	Anzahl der Schwestern
r, r_s	Bildungsrendite, Diskontfaktor, Gegenwartsvorliebe
r_t	Rendite von firmenspezifischem Kapital
σ	Standardabweichung in der Grundgesamtheit
s	Dauer der allgemeinen Bildung
U	Nutzen
V_s	Abdiskontiertes Lebensarbeitseinkommen
$verhei_i$	Individuum i ist verheiratet
w_i^o	Beobachtbarer Lohn von Individuum i
w_i^r	Unbeobachtbarer Lohn, Reservationslohn von Individuum i
$weibl_i$	Individuum i ist weiblich
x	Dauer der firmenspezifischen Bildung, Berufserfahrung
Y_t	Einkommen in Periode t
ZKH_i	Zahl der Kinder im Haushalt von Individuum i

1 Einleitung

In dieser Arbeit werden nationale und regionale Bildungsrenditen theoretisch wie empirisch analysiert. Die Motivation zur Behandlung sowie eine Einführung in dieses Thema werden im ersten Unterkapitel dieser Einleitung erläutert. Der zweite Teil zeigt die Struktur der Arbeit.

1.1 Fragestellung

„(Aus)Bildung ist der beste Reiseproviand für die Reise zum hohen Alter“ formulierte bereits Aristoteles, um die hohe Bedeutung von Wissen für das Individuum deutlich zu machen. Dabei wird unter Bildung die Aneignung bzw. Weiterentwicklung von Fähigkeiten und Fertigkeiten verstanden. Es können nun viele Formen von Wissen unterschieden werden. In dieser Arbeit wird im besonderen Maße das für den Arbeitsmarkt qualifizierende Wissen untersucht, welches im Allgemeinen in generelles und firmenspezifisches Humankapital zu unterscheiden ist. Für beide Formen wird von einer Produktivitätssteigerung des Individuums durch dessen Aneignung ausgegangen. Schon vor ca. 250 Jahren stellte Benjamin Franklin fest, dass „eine Investition in Wissen [...] immer noch die besten Zinsen“ bringt. Diese Investitionsbetrachtung in die individuelle Bildung wird in dieser Arbeit aufgegriffen. Dabei ist davon auszugehen, dass die aktive Wissensaneignung mit Kosten verbunden ist, die sich über kurz oder lang amortisieren müssen. Es ist anzunehmen, dass durch eine höhere Bildung auch das Einkommen steigt, es gilt jedoch, die Höhe dieser Steigerung zu thematisieren. Konkret wird die jährlich zusätzliche Entlohnung durch ein weiteres Bildungsjahr untersucht. Diese sogenannte Bildungsrendite definiert Weber (2003), S. 406, als „inwieweit der zu erwartende Nutzen einer Bildungsinvestition die Kosten zu kompensieren vermag.“ Bei der Investition in Bildung wird davon ausgegangen, dass in der ersten Lebensphase, der Bildungsphase, Kosten in Form eines nicht generierten Einkommens entstehen, die in der zweiten Lebensphase, der Arbeitsphase, durch ein höheres Einkommen (über)kompensiert werden. Ziel dieser Arbeit ist es, die

1 Einleitung

Höhe des reinen Effekts von generellem Humankapital auf das Einkommen zu ermitteln. Die besondere Herausforderung dabei ergibt sich aus der großen Verschiedenartigkeit der Individuen. Eine Korrektur um individuelle Unterschiede, zum Beispiel in der Motivation, ist notwendig. Diese Korrekturen werden über die dargestellten Verfahren näherungsweise vorgenommen. Es gilt als Ziel dieser Arbeit, eine Intervallspanne für die Bildungsrenditen zu errechnen. Hierfür werden diverse statistische Verfahren und Modellspezifikationen diskutiert.

Um der Heterogenität der Individuen gerecht zu werden, ist neben der gesamtdeutschen Analyse auch die Darstellung für verschiedene Charakteristika vorgesehen. Dabei werden verschiedene Alterscluster sowie der Unterschied zwischen öffentlichem Dienst und der Privatwirtschaft betrachtet. Zudem erfolgt eine Ermittlung der Bildungsrendite separat für Teil- und Vollerwerbstätige. Doch gilt es nicht nur die Verschiedenartigkeit der Individuen an sich zu betrachten, unterschiedliche regionale Aspekte können ebenso beeinflussend wirken. Deutschland ist ein sehr heterogenes Land. Diverse Studien diskutieren im Besonderen regionale Disparitäten. So werden von Berthold und Müller (2010) die Angleichungsprozesse über die Jahre in Bezug auf die Lebensverhältnisse in den Regionen Deutschlands thematisiert. Auch Tarazona (2010) betrachtet die interregionalen Unterschiede und zeigt in diesem Zusammenhang verschiedene Beschäftigungsentwicklungen auf.

Der Bezug zur Heterogenität von Bildungsrenditen wird in einer aktuellen internationalen Studie von Mendolicchio und Rhein (2012) hergestellt. Sie berechnen unterschiedliche Höhen der jährlichen Verzinsung von Humankapital im internationalen Vergleich der westeuropäischen Staaten. Diese aktuelle Studie und der Blick auf diverse regionale Disparitäten wirft für den innerdeutschen Vergleich die Frage der Konstanz in den regionalen Bildungsrenditen auf. Bildungspolitik wird im föderalistischen deutschen Staat bundesländerintern entschieden. Daher und mit Blick auf die heterogene Struktur in Deutschland, sind durchaus regional verschiedene Bildungsrenditen zu erwarten. In dieser Arbeit wird daher die Frage diskutiert, ob Deutschland, aufgrund vieler regionaler Disparitäten, auch in der Bewertung von allgemeinem Humankapital Unterschiede aufweist, und in welchem Bundesland eine besonders hohe bzw. besonders niedrige Verzinsung zu erwarten ist.

1.2 Aufbau der Arbeit

Diese Arbeit diskutiert zunächst im 2. Kapitel das theoretische Fundament zur Ermittlung des Einflusses von Humankapital auf das Einkommen. Dabei ist es das vorrangige Ziel, eine Lohngleichung zu entwickeln, die diesen Zusammenhang abbildet. Es werden zwei Herangehensweisen nach Jacob Mincer ausführlich dargestellt. Das 3. Kapitel zeigt den aktuellen Stand der Forschung in der empirischen Schätzung der Bildungsrenditen in Deutschland. Dabei werden einige Studien gegenübergestellt und in Verbindung mit den jeweils verwendeten Verfahren diskutiert.

Nach dieser Einführung folgt in Abschnitt 4 die Darstellung des Datensatzes. Im Besonderen wird dabei auf seine Zusammensetzung eingegangen und eine Prüfung der Repräsentativität vorgenommen. Der Datensatz wird mit dem Mikrozensus deskriptiv verglichen. Es folgt eine genaue Beschreibung der, für die weitere Analyse wesentlichen Variablen.

In Kapitel 5 werden zunächst die verschiedenen statistischen Verfahren zur Schätzung der Bildungsrendite diskutiert. Dabei wird auf das OLS-Verfahren und die zugehörigen Voraussetzungen, auf das 2SLS-, das 3SLS- sowie das Heckman-Verfahren eingegangen. Zunächst werden die drei erstgenannten Verfahren auf den beschriebenen Datensatz mit Erwerbstätigen angewandt. Über den Vergleich der gesamtdeutschen Resultate für verschiedene Modellspezifikationen werden sinnvolle Verfahren zur Schätzung der Bildungsrendite identifiziert. Dieses gilt es im nächsten Schritt anzuwenden.

Zunächst wird eine Separierung nach Altersgruppen vorgenommen. Im Speziellen sind die Cluster 30–40, 41–50 und 51–60-Jährige zu untersuchen. Hierbei wird unterstellt, dass sich die Bildungsrenditen, unabhängig von der Berufserfahrung mit verschiedenem Alter unterscheiden. Weiterhin wird angenommen, dass sich die Renditen zwischen Erwerbstätigen im öffentlichen Dienst und der Privatwirtschaft unterscheiden. Es erfolgt eine weitere Separierung des Datensatzes und die wiederum getrennte Berechnung der Bildungsrenditen. Die letzte Merkmalsunterscheidung wird in Bezug auf Voll- und Teilzeitbeschäftigung vorgenommen. Auch an dieser Stelle ergeben sich separat ermittelte Bildungsrenditen. Damit wird geprüft, ob für die einzelnen Unterscheidungsmerkmale in der erwerbstätigen Bevölkerung tatsächlich verschiedene Bildungsrenditen zu schätzen sind.

Es folgt eine Erweiterung der Aussagekraft der Ergebnisse auch auf Erwerbslose. Zu diesem Zweck wird der Datensatz um die ALG-I-Empfänger erweitert und mit dem 3SLS-Verfahren geschätzt. Mit dem Heckman-Verfahren ist es im Weiteren möglich alle Erwerbspersonen in die Betrachtung mit einzubeziehen. Deshalb wird der Datensatz noch-

1 Einleitung

mals erweitert. Zunächst erfolgt eine Repräsentativitätsprüfung dieses Datensatzes, um diesen für eine nationale Analyse anzuwenden.

Auf die regionalen Unterschiede wird in Kapitel 6 eingegangen und bildet den Hauptteil der Arbeit. Für einen ersten Eindruck erfolgt zunächst eine Ost-West-Betrachtung, die dann auf Bundesländerebene ausgebaut wird. Hierbei werden verschiedene Regressionsmodelle zur Schätzung regionaler Bildungsrenditen entwickelt und die Ergebnisse miteinander verglichen. Dabei wird zum einen eine Modellvariante mit separiertem, zum anderen eine mit einem gesamtdeutschen Datensatz untersucht. Zusätzlich findet eine regionale Betrachtung mit allen Erwerbspersonen statt. Auch hierfür wird die Heckman-Methode verwendet. Sie wird ebenfalls sowohl für einen separierten, als auch für einen gesamtdeutschen Datensatz angewendet.

Es folgt ein vergleichendes Teilkapitel, welches die Ergebnisse der verschiedenen Schätzungen der regionalen Bildungsrenditen gegenüberstellt. Auf Basis dieser Ergebnisse werden mögliche Ursachen und Auswirkungen der geschätzten regionalen Bildungsrenditen diskutiert. Diese erfolgt über einen bivariaten Vergleich mit einigen regional verschiedenen Merkmalen. Die Arbeit schließt mit einem zusammenfassenden Fazit sowie einem kurzen Ausblick, in dem weiterer Forschungsbedarf aufgezeigt wird.

2 Theoretische Fundierung

Becker (1994 (3. Auflage)) unterscheidet zwischen generellem und firmenspezifischen Humankapital (HC). Die Generierung dieses Wissens erfolgt in zwei getrennten Lebensphasen. Zusammen bilden sie die produktive Lebensphase. Im ersten Teil wird allgemeines Humankapital generiert. Darunter wird die schulische Ausbildung, die berufsqualifizierende schulische Ausbildung sowie ein Studium verstanden. Das firmenspezifische Humankapital wird im zweiten Teil der produktiven Lebensphase erzeugt. Dabei wird Wissen durch berufsbezogene Weiterbildungen oder direkt am Arbeitsplatz generiert. In Abbildung 2.1 sind beide Lebensphasen anhand eines Zeitstrahls verdeutlicht.

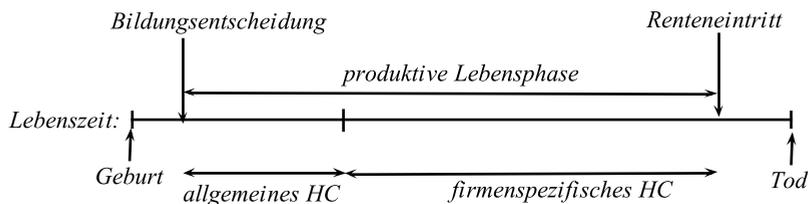


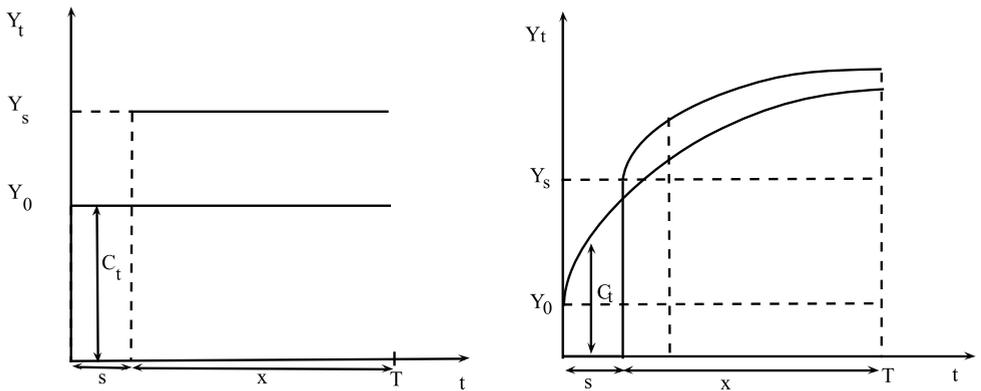
Abbildung 2.1: Lebenszyklus bei Unterscheidung in generelles und firmenspezifisches Humankapital, eigene Darstellung.

Beide Phasen werden annahmegemäß im Modell voneinander getrennt. Phase 1 ist eine reine Bildungsphase. Dies bedeutet, dass die gesamte zur Verfügung stehende Zeit zur Akkumulation von allgemeinem Humankapital aufgewandt und demnach in dieser Zeit kein Einkommen entsteht. Durch diesen Zuwachs an Bildung, bei längerer Dauer der ersten Phase, kann in der zweiten Phase ein höheres Einkommen erzielt werden. Demzufolge steigen die potentiellen Einkommen mit dem Bildungsniveau. Je länger sich das Individuum in Phase 1 bildet, umso höher sind damit die entgangenen Einkommen, bei weiterer Erzeugung von Bildung. Ab Phase 2 kann das Individuum zu einem firmenspezifischen Humankapital akkumulieren, zum anderen ist die Phase geprägt durch die Generierung von Einkommen.

2 Theoretische Fundierung

Es ist davon auszugehen, dass je länger sich das Individuum bildet, umso höher ist das Einkommen in Phase 2. Allerdings wird durch eine längere Bildungsphase ebenso die Phase der Einkommensgenerierung verkürzt. Weiterhin ist die Zeit, in der dem Individuum kein Einkommen zur Verfügung steht, bei langer allgemeiner Bildungsphase auch lang. Es gilt daher bei der Entscheidung über die Investition in Bildung die Vor- und Nachteile gegenüberzustellen.

Zwei Bildungsalternativen. Betrachtet man zunächst ein Individuum, welches zwischen zwei Bildungsalternativen entscheidet und kein firmenspezifisches Humankapital generiert. Es ergeben sich zwei mögliche Zahlungsströme, wie in Abbildung 2.2.a. dargestellt. Verglichen werden hierbei die Einkommensströme nach *null* bzw. *s* Bildungsjahren. Für die erste Alternative gilt, dass das Individuum sofort mit der zweiten, der einkommensgenerierenden Phase startet. Bei einem Niveau von null² Bildungsjahren kann ein Einkommen in Höhe von Y_0 generiert werden. Bildet sich die Person nun *s* Jahre, so kann sie nach dieser Zeit ein deutlich höheres Einkommen Y_s erzielen, verzichtet aber die ersten Jahre auf das geringere Einkommen Y_0 . Dieses potentielle, aber nicht generierte Einkommen sind Opportunitätskosten und annahmegemäß die einzigen Kosten der Bildung.



(a) Einkommensstrom bei ausschließlicher Betrachtung des allgemeinen Humankapitals. (b) Einkommensstrom bei Berücksichtigung von allgemeinem und firmenspezifischem Humankapital.

Abbildung 2.2: Einkommensstrom für hohe und niedrige Qualifizierung; in Anlehnung an Borjas (1996), S. 224. und Mincer (1974), S. 17.

²Null muss nicht zwangsläufig keine Bildung heißen. Man könnte hierbei auch von einem grundsätzlichen Bildungsstand ausgehen und dann null zusätzliche Bildungsjahre annehmen.

2.1 Der Einfluss von allgemeinem Humankapital – Das Grundmodell nach Jacob Mincer

Berücksichtigt man in einem zweiten Schritt neben dem Einfluss von allgemeinem Humankapital auch das firmenspezifische, so ergibt sich die Darstellung in 2.2.b. Auch hierbei werden zwei Bildungsalternativen miteinander verglichen. Es wird angenommen, dass firmenspezifisches Humankapital einen positiv abnehmenden Grenzertrag³ hat. Das bedeutet, dass das erste Jahr einkommenssteigernder wirkt als das letzte. Der Unterschied im Vergleich zur vorhergehenden Darstellung ist, dass nach Beendigung der allgemeinbildenden Phase das Einkommen über die Arbeitszeit nicht konstant ist, sondern mit zunehmenden Berufsjahren steigt. Demzufolge entfällt für das sich bildende Wirtschaftssubjekt zum einen das Einkommen während der Bildungsphase, zum anderen ist die Dauer zu Generierung einer Einkommenssteigerung durch die Akkumulation von firmenspezifischem Kapital kürzer. Ziel dieser Arbeit ist es, näherungsweise eine mittlere Rendite für generelles Humankapital zu ermitteln. Hierfür soll zunächst eine theoretische Diskussion erfolgen, die den Zusammenhang zwischen Bildung und Einkommen erklärt.

2.1 Der Einfluss von allgemeinem Humankapital – Das Grundmodell nach Jacob Mincer

In diesem Abschnitt wird ein einfaches Modell von Jacob Mincer zur Darstellung des Zusammenhangs zwischen Einkommen und Bildung diskutiert. Im ersten Schritt wird dabei die Länge der produktiven Lebensphase konstant angesehen. Im zweiten Schritt wird diese Annahme gelockert, sodass dann die Länge der einkommensgenerierenden Phase als konstant zu verstehen ist. Am Schluss folgt eine Kritik an den Modellen.

2.1.1 Grundmodell

Aufbauend auf die Theorie von Becker (1994 (3. Auflage)) schuf Mincer (1974) eine theoretische Herleitung für den Zusammenhang zwischen Bildung und Einkommen. Dabei werden in dem Grundmodell zunächst lediglich zwei Bildungsalternativen mit den zugehörigen Einkommensströmen verglichen. Gegenübergestellt werden dabei die diskontier-

³Die theoretische Herleitung folgt in Abschnitt 2.2 sowie die empirische Diskussion in Abschnitt 4.

2 Theoretische Fundierung

ten Lebensarbeitseinkommen zum Zeitpunkt der Bildungsentscheidung. Der Einkommensstrom wird als diskreter⁴ Zahlungsstrom pro Bildungsniveau interpretiert.

Verglichen wird der Einkommensstrom nach s und nach $(s - d)$ Bildungsjahren, wobei angenommen wird, dass $0 < d \leq s$ ist. Die erste der beiden Bildungsalternativen symbolisiert ein höheres Humankapital. Daher wird das Einkommen Y_s größer sein als jenes der Vergleichsalternative. Bei letztgenannter hat das Wirtschaftssubjekt eine um d Jahre verringerte Bildungsdauer und kann daraus nur einen geringeren periodischen Einkommensstrom in Höhe von Y_{s-d} erzielen. Der Vorteil hierbei hingegen ist, dass die einkommensgenerierende Zeit früher startet und länger währt. In 2.3 sind beide Bildungsalternativen gegenübergestellt.

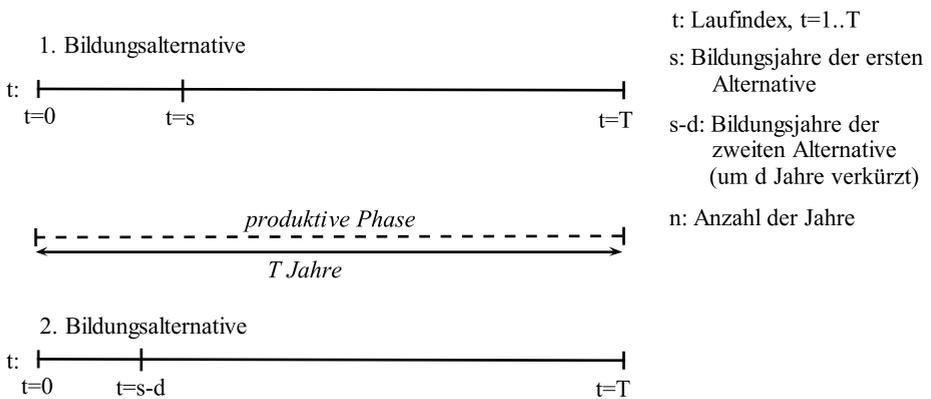


Abbildung 2.3: Lebenszyklus, eigene Darstellung.

Mincer (1974) setzt folgende Grundannahmen:

- A1** Die Höhe des Einkommens wird ausschließlich durch das allgemeine Humankapital definiert. Alle weiteren Einflüsse, z. B. auch jener von firmensepezifischem Kapital werden nicht berücksichtigt.
- A2** In der allgemeinbildenden Phase wird kein Einkommen erzielt. Die hier entstehenden Opportunitätskosten in Form von entgangenem Einkommen sind die einzigen Kosten der Bildung. Weiterhin wird durch zunehmende Bildung von allgemeinem Humankapital die Zeit der Einkommensgenerierung verkürzt.

⁴Mincer wendet sein Modell auch für eine stetige Reihe an. Da Einkommen üblicherweise diskret ausbezahlt werden und das stetige Modell keinen Informationszuwachs liefert, ist es ausreichend an dieser Stelle lediglich das diskrete Modells zu zeigen.

A3 Die Verzinsung eines zusätzlichen Bildungsjahres ist konstant.

A4 Das rationale handelnde Individuum entscheidet sich immer dann für Bildung, wenn der Barwert der Lebensarbeitseinkommen der höheren Bildungsalternative größer dem Barwert der geringeren Alternative mit geringerem Bildungsniveau ist. Das Individuum ist indifferent, wenn gilt, dass die Barwerte gleich sind und demnach $V_s = V_{s-d}$.

Zunächst werden die diskontieren Lebensarbeitseinkommen der beiden Alternativen getrennt voneinander diskutiert. So ist für s Jahre Bildung, ein kontinuierlicher periodischer Einkommensstrom in Höhe von Y_s zu erwarten. Die Generierung startet im ersten Jahr nach der allgemeinbildenden Phase und endet mit dem Renteneintrittsalter T . Um den Barwert zu ermitteln, müssen die zukünftigen Einkommen auf den Entscheidungszeitpunkt diskontiert werden. Dabei wird das zukünftige Einkommen mit der Zeitpräferenzrate r diskontiert, wobei jedes Jahr, welches weiter in der Zukunft liegt, stärker reduziert wird. Es ergibt sich

$$V_s = Y_s \sum_{t=s+1}^T \frac{1}{(1+r)^t}, \quad (2.1)$$

Dieses Gleichung kann über die Lösungsformel des Rentenbarwertfaktors gelöst werden und damit ergibt sich Gleichung (2.1) zu

$$V_s = Y_s \frac{(1+r)^T - (1+r)^s}{r(1+r)^{s+T}}. \quad (2.2)$$

Die gleiche Herangehensweise gilt auch für die zweite Bildungsalternative mit $(s-d)$ Bildungsjahren. Hierbei startet die einkommensgenerierende Phase zum Zeitpunkt $(s-d+1)$ und endet – wie zuvor – nach T Jahren mit dem Ende der produktiven Phase. Dabei wird ein geringerer Einkommensstrom in Höhe von Y_{s-d} generiert, der unter Berücksichtigung der Gegenwartsvorliebe zu diskontieren ist.

$$V_{s-d} = Y_{s-d} \sum_{t=s-d+1}^T \frac{1}{(1+r)^t}. \quad (2.3)$$

Ebenso wie für V_s wird auch Gleichung (2.3) gelöst. Es ergibt sich

$$V_{s-d} = Y_{s-d} \frac{(1+r)^{T+d} - (1+r)^s}{r(1+r)^{s+T}}. \quad (2.4)$$

Die Investition in Bildung wird bei diesem Vergleich solange zugunsten von Humankapital ausfallen, so lange der Barwert nach s Jahren größer ist als jener nach $(s-d)$ Jahren. Der kritische Wert ergibt sich demnach genau, wenn beide Alternativen die gleichen Barwerte erzeugen (siehe A4). Daher gilt

$$V_s = V_{s-d}$$

$$Y_s \frac{(1+r)^T - (1+r)^s}{r(1+r)^{s+T}} = Y_{s-d} \frac{(1+r)^{T+d} - (1+r)^s}{r(1+r)^{s+T}}. \quad (2.5)$$

Durch Umformungen erhält man

$$\frac{Y_s}{Y_{s-d}} = \frac{(1+r)^{T+d} - (1+r)^s}{(1+r)^T - (1+r)^s}. \quad (2.6)$$

Vergleicht man nun die Alternative s mit dem niedrigsten Bildungsniveau, so muss gelten, dass $s = d$ ist. Unter Berücksichtigung dieser Bildungsalternativen ergibt sich

$$Y_s = Y_0 \frac{(1+r)^{n+s} - (1+r)^s}{(1+r)^n - (1+r)^s}. \quad (2.7)$$

Dieses Ergebnis zeigt eine mathematische Lösung, ist allerdings wenig aussagekräftig für eine weitere Diskussion. Demnach soll die Annahme der konstanten produktiven Lebensphase aufgehoben werden und stattdessen, die einkommensgenerierende Phase als konstant angenommen werden.

2.1.2 Grundmodell mit konstanter Zahl an Berufsjahren

Das betrachtete Modell wird leicht modifiziert. Dabei wird nicht mehr davon ausgegangen, dass jedes Individuum zum gleichen Zeitpunkt die produktive Phase verlässt und damit ins Rentenalter eintritt. Es wird vielmehr angenommen, dass die einkommensgenerierenden Jahre unabhängig von der Bildungsdauer konstant sind. Der Vorteil einer früheren Einkommensgenerierung bei kurzer Bildungsdauer bleibt bestehen; der Vorteil der längeren Zeit

2.1 Der Einfluss von allgemeinem Humankapital –Das Grundmodell nach Jacob Mincer

um Einkommen zu erzielen, wird jedoch negiert. Die Auswirkungen dieser zusätzlichen Annahme sind jedoch als gering einzuschätzen, da es sich um eine Reduktion der *letzten* Berufsjahre handelt. Bei der Ermittlung des Barwertes werden alle Einkommensströme je ferner die Generierung in der Zukunft liegt, zunehmend stärker diskontiert. Demzufolge ist der Einfluss den die hier unberücksichtigten Jahre auf den Barwert des Lebensarbeitsentkommens hätten gering. Die Veränderungen des Modells werden in Abbildung 2.4 verdeutlicht.

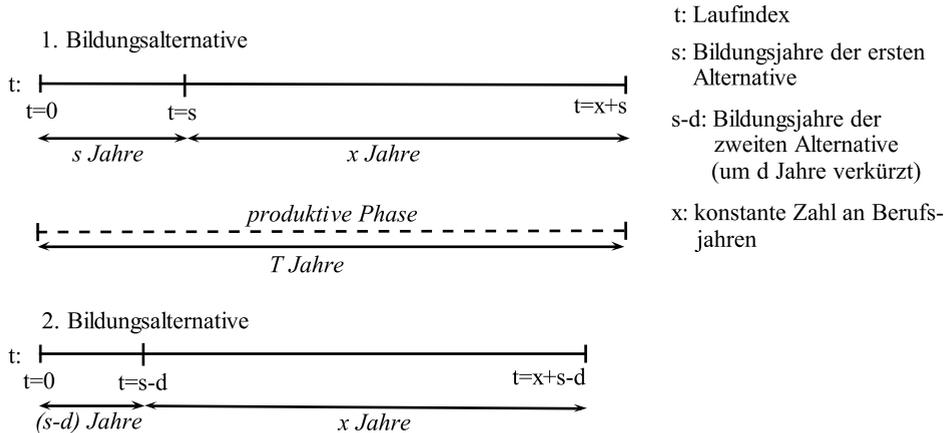


Abbildung 2.4: Lebenszyklus für eine konstante Zahl an Berufsjahren, eigene Darstellung.

Abgesehen von der Änderungen der konstant angenommen Größe wird das Modell nicht weiter modifiziert. Die dargestellten Grundannahmen bleiben bestehen. Auch die Überlegung, zwei Bildungsalternativen gegenüberzustellen und die jeweiligen Barwerte zu vergleichen, bleibt.

Zur Darstellung des diskontierten Lebensarbeitsentkommens nach s Bildungsjahren ändert sich, durch die neue Modellannahme lediglich die Obergrenze der Einkommensgenerierung. Es wird nun, statt zuvor von T Jahren auszugehen, vom Zeitpunkt $(x+s)$ als Ende der produktiven Zeit ausgegangen. Dabei ist x die, unabhängig von der gewählten Bildungsalternative, konstante Zahl an Jahren, in denen Einkommen erzielt wird. Der Barwert für dieses Bildungsniveau kann demnach berechnet werden als

$$V_s = Y_s \sum_{t=s+1}^{x+s} \frac{1}{(1+r)^t} \quad (2.8)$$

2 Theoretische Fundierung

Löst man diese Gleichung wie zuvor, so erhält man:

$$V_s = Y_s \left[\frac{(1+r)^x - 1}{r(1+r)^{x+s}} \right]. \quad (2.9)$$

Auch für die zweite Bildungsalternative nach $(s-d)$ Jahren ergibt sich die gleiche Änderung. Nun endet die produktive Phase bereits nach $(x+s-d)$ Jahren. Damit ist die gesamte produktive Lebensphase bei dieser Alternative um d Jahre kürzer. Der Barwert kann damit berechnet werden mit

$$V_{s-d} = Y_{s-d} \sum_{t=s-d+1}^{x+s-d} \frac{1}{(1+r)^t}.$$

Die Lösung der Summe ergibt

$$V_{s-d} = Y_{s-d}(1+r)^d \left[\frac{(1+r)^x - 1}{r(1+r)^{x+s}} \right]. \quad (2.10)$$

Auch an dieser Stelle werden beide diskontieren Lebensarbeitseinkommen einander gegenüber gestellt. Der kritische Wert ist dort erreicht, wo beide Alternativen die gleichen Barwerte aufweisen. Daher werden wieder die Alternativen V_s aus Gleichung (2.9) und V_{s-d} aus Gleichung (2.10) gleichgesetzt. Es ergibt sich

$$V_s = V_{s-d}.$$

$$Y_s \left[\frac{(1+r)^x - 1}{r(1+r)^{x+s}} \right] = Y_{s-d}(1+r)^d \left[\frac{(1+r)^x - 1}{r(1+r)^{x+s}} \right]. \quad (2.11)$$

Durch Kürzen lässt sich Gleichung (2.11) umschreiben zu

$$\frac{Y_s}{Y_{s-d}} = (1+r)^d. \quad (2.12)$$

Vergleicht man nun wieder die Bildungsdauer s mit einem weiteren spezifischen Bildungsniveau, nämlich jenem nach null zusätzlichen Bildungsjahren, so gilt $(s=d)$ und wirkt auf Gleichung (2.12) so, dass nun der Zusammenhang geschrieben werden kann als

$$Y_s = Y_0(1+r)^s.$$

2.1 Der Einfluss von allgemeinem Humankapital –Das Grundmodell nach Jacob Mincer

Logarithmiert man nun beide Seiten und wendet die Logarithmengesetze⁵ an, so erhält man die bekannte Einkommensgleichung von Jacob Mincer:

$$\ln Y_s \approx \ln Y_0 + rs. \quad (2.13)$$

Demzufolge wird das logarithmierte Einkommen nach s Bildungsjahren zum einen aus dem logarithmierten Einkommen welches unabhängig von der Bildungsdauer generiert wird, bestimmt. Zum anderen wirkt die Bildungsdauer an sich einkommenssteigernd. Mit welcher Rate das Einkommen nun durch das höhere Bildungsniveau steigt, zeigt die Rendite r . Dabei ist r anfänglich als Diskontfaktor in das Modell integriert worden. Hierbei gilt, je höher die Zeitpräferenzrate ist, desto stärker wird zukünftiges Einkommen diskontiert. Demzufolge werden die in der Zukunft liegenden Einkommensströme in diesem Fall deutlich stärker reduziert als es bei einer geringeren Gegenwartsvorliebe der Fall wäre. Damit das Individuum in Bildung investiert, müsste demnach die Rendite bei hoher Zeitpräferenzrate ebenfalls hoch sein. Gleicher Zusammenhang gilt bei geringer Zeitpräferenzrate. Somit kann ein Zusammenhang zwischen der Zeitpräferenzrate und der Bildungsrendite hergestellt werden.

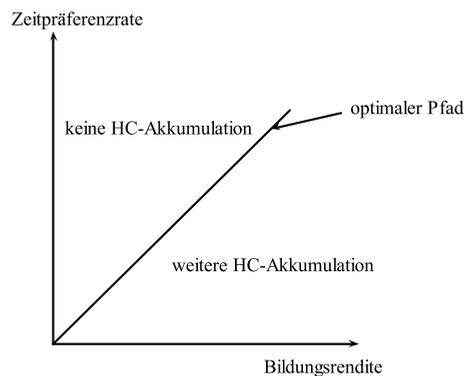


Abbildung 2.5: Zusammenhang zwischen der Bildungsrendite und der Zeitpräferenzrate und die Auswirkung auf die Bildungsentscheidung, eigene Darstellung.

In Abbildung 2.5 wird dieser Zusammenhang überblicksartig dargestellt. Dabei wird grob von einem linearen Zusammenhang ausgegangen. Es ist nicht sicher, ob dieser Verlauf für alle möglichen Zeitpräferenzraten unterstellt werden kann, soll aber an dieser Stelle

⁵Allgemein gilt für kleine Werte von x approximativ, dass $\ln(1+x) \sim x$. Damit gilt hier $\ln(1+r) \sim r$.

2 Theoretische Fundierung

vereinfachend angenommen werden. Es wird jedoch deutlich, dass beide Größen gleichgerichtet zusammenhängen und demnach eine hohe Zeitpräferenzrate mit höherer Rendite und eine niedrige Zeitpräferenzrate mit niedrigerer Rendite einhergeht.

Die optimale Bildungsentscheidung kann nun auch als Verhältnis zwischen beiden Größen beschrieben werden:

$$g = \frac{\text{Zeitpräferenzrate}}{\text{Bildungsrendite}}.$$

So gilt, wenn

- $g < 1$, dass die Gegenwartsvorliebe kleiner als die Verzinsung ist. Demnach wird der aktuelle Verzicht auf Einkommen durch eine zukünftig höhere Verzinsung ausgeglichen. Das Individuum wird weiter in Bildung investieren.
- $g = 1$, dass sich Gegenwartsvorliebe und Bildungsrendite entsprechen. An dieser Stelle ist das Individuum indifferent bei der Entscheidung über eine Investition in weitere Bildung. Hier ergibt sich der optimale Bildungspfad, wie dargestellt in Abbildung 2.5.
- $g > 1$, dass die Gegenwartsvorliebe größer als die Verzinsung ist. Es lohnt sich daher nicht mehr für das Individuum in Bildung zu investieren.

Anhand dieses Modells konnte ein Beziehung zwischen dem Einkommen und der Bildungsdauer hergestellt werden. Mincer schuf mit dem Modell eine empirisch schätzbare, theoretisch fundierte Gleichung zur Quantifizierung dieses Zusammenhangs.

2.1.3 Kritik

Mit diesem Modell schafft Jacob Mincer aufbauend auf die Humankapitaltheorie von Gary Becker ein theoretisches Fundament für den Zusammenhang zwischen Humankapital und dem Einkommen. Die mikroökonomische Fundierung der individuellen Bildungsentscheidung in Zusammenhang mit der Rendite und der Zeitpräferenzrate des Grundmodells schafft es einige Aspekte dieser vielfältigen Entscheidung zusammenzufassen. Generell muss bemängelt werden, dass es deutlich mehr Einflußfaktoren für das Einkommen gibt. So wird es beispielsweise beeinflusst von intrinsischer Motivation, der Wirtschaftslage, der Region oder Ähnlichem. Gleiches gilt für die Bildungsdauer. Auch sie ergibt sich nicht nur aus dem Zusammenspiel von Rendite und Gegenwartsvorliebe.

2.1 Der Einfluss von allgemeinem Humankapital –Das Grundmodell nach Jacob Mincer

Mincer geht davon aus, dass alle Individuen zu einem bestimmten Zeitpunkt ($t=0$) ihre Bildungsentscheidung treffen. Irrelevant dabei ist, ob dieser Zeitpunkt beim Startpunkt der Bildungsakkumulation oder bei einem grundsätzlichen Bildungsniveau liegt, welches jedes Individuum⁶ erreicht. Dies ist kritikwürdig, da sich die Bildungsentscheidung realistischerweise in einem evolutorischen Prozess ergibt. Schritt für Schritt wird jedes Individuum seine aktuellen Chancen und Risiken austarieren und entsprechend seiner Erwartungen die weitere Bildungsakkumulation gestalten.

Die Annahme, dass jedes Individuum sein Lebensarbeitseinkommen pro Bildungsalternative kennt und zum Entscheidungsmoment berechnet, kann nicht als gegeben angesehen werden. Zudem ist die Annahme, dass jedes Wirtschaftssubjekt als homo oeconomicus und demnach reiner Einkommensmaximierer zu verstehen ist, nicht gegeben. Die Bildungskosten werden rein auf die Opportunitätskosten aus entgangenem Einkommen reduziert. Lebenshaltungskosten oder andere Bildungskosten, wie Studiengebühren, sind nicht berücksichtigt. Tatsächlich entstehen dem sich weiterbildenden Individuum zusätzlich auch nichtmonetär messbare Kosten. Auch diese finden keine Berücksichtigung.

In Mincers Modell wird Bildung als homogenes Gut verstanden. Zum Beispiel zählt ein Studium, egal welches Fach studiert wird, als gleichermaßen produktivitätssteigernd. Zudem wird unterstellt, dass jedes weitere Bildungsjahr in gleicher Weise einkommenssteigernd wirkt, wie das vorherige.

Aufgrund der Unlösbarkeit des in Abschnitt 2.1 beschriebenen Modells wird die Annahme eingeführt, dass die Zahl der Berufsjahre eines jeden Individuums konstant sind. Wie bereits in Abschnitt 2.2 beschrieben, dürfte jedoch die Veränderung durch diese Annahme im Modell recht gering ausfallen. Sie betrifft die *letzten* Berufsjahre, die Einkommen dieser Berufsjahre werden deutlich diskontiert, sodass der Einfluss auf den Barwert des Lebensarbeitseinkommens eher von marginaler Natur ist.

In der bisherigen Betrachtung werden Einkommensdifferenzen lediglich durch ein verschiedenes Bildungsniveau erklärt. Eine Einkommenssteigerung verursacht durch unterschiedliches firmenspezifisches Humankapital ist in diesem Modell nicht berücksichtigt. Nachfolgend wird daher eine Erweiterung um diesen Einfluss diskutiert.

⁶In Abschnitt 4.2 wird dieses näher beschrieben. Angenommen wird, dass jedes Individuum ein Mindestbildungsniveau von 7 Bildungsjahren erreicht.

2.2 Einfluss von allgemeinem und firmenspezifischem Humankapital

In diesem Abschnitt wird eine weitere Darstellung für die Herleitung der Mincer-Gleichung gezeigt. Dabei werden beide Arten von Humankapital in die Betrachtung einbezogen. Außerdem erfolgt die Herleitung nicht über den Vergleich der Barwerte zweier Bildungsalternativen, sondern es wird das aktuelle Einkommen, welches sich aus der Bildung der Vorperioden zusammensetzt, erklärt. Nach der theoretischen Herleitung folgt eine kritische Auseinandersetzung mit diesem Modell.

2.2.1 Erweitertes Modell von Jacob Mincer

Zur Erweiterung der Lohngleichung um die Jahre der Berufserfahrung wird wiederum ein Ansatz von Jacob Mincer⁷ genutzt. Aus Lösbarkeitsgründen kann jedoch das in Abschnitt 2.1.2 dargestellte Modell nicht einfach um die Berufserfahrung erweitert werden.

Dieses Modell zielt nun auf eine Erklärung des aktuellen Einkommens, nicht mehr auf das diskontierte Lebensarbeitseinkommen. Es baut auf folgenden Grundannahmen auf:

A1 Die produktive Lebensphase wird auf s Bildungsjahre sowie x Berufsjahre aufgeteilt. Im Modell wird davon ausgegangen, dass es keine Phasen der Nichtbeschäftigung (Arbeitslosigkeit, Elternzeit oder Ähnliches) gibt.

A2 Der individuelle Humankapitalbestand HC kann über die Zeit t nur zunehmen. Es gilt daher, dass

$$\frac{\partial HC}{\partial t} > 0.$$

A3 Während der ersten Periode s wird die komplette Zeit zur Akkumulation von Humankapital verwendet und daher kein Einkommen generiert. Es ergeben sich in dieser Phase keine direkten Kosten, sondern einzig Opportunitätskosten in Form von entgangenem Einkommen.

A4 Allgemeines Humankapital verzinst sich unabhängig vom Bildungsniveau mit einer konstanten Rate r_s . Damit gilt in der einkommensgenerierenden Phase, dass für alle Zeitpunkte, wenn $t < s$, die allgemeine Rendite r_t konstant ist. Somit ergibt sich $r_t = r_s = \text{const.}$, für alle $t < s$.

⁷Dieser Abschnitt basiert teilweise auf den Artikeln von Mincer (1970) und Heckman et al. (2006).

A5 Auch die Generierung von firmenspezifischem Wissen ist mit Opportunitätskosten verbunden. Dieses Humankapital verzinst sich ebenso wie das allgemeine mit einer konstanten Rate. Jedoch sind beide Raten unabhängig voneinander. Es gilt demnach für $allet \geq s$, dass $r_t = r_x = const.$

Anders als in den Modellen zuvor wird nun nicht das diskontierte Lebensarbeitseinkommen betrachtet, sondern der Fokus auf die Ermittlung der Zusammensetzung des aktuellen tatsächlich erzielten Einkommens gelegt. Ausgangspunkt ist die Berücksichtigung einer Investitionsquote. Sie ergibt sich aus der Relation zwischen Kosten und potentiellm Einkommen. Es gilt

$$k_t = \frac{C_t}{E_t}, \quad (2.14)$$

wobei k_t für die Investitionsquote in Periode t steht, C_t zeigt die durch weitere Bildung in Periode t entstehenden Opportunitätskosten und E_t steht für das potentielle Einkommen. Für die Zeit der Erzeugung von allgemeinbildendem Humankapital gilt, dass das komplette potentielle Einkommen dieser Periode reinvestiert wird. Damit entsprechen sich in dieser Phase potentielles Einkommen sowie Opportunitätskosten. Es gilt daher

$$k_t = 1, \text{ für } t < s.$$

Die Investitionsquote liegt demnach in dieser Zeit bei 100 %. Das tatsächliche Einkommen ist dann annahmegemäß null. Es berechnet sich aus:

$$Y_t = E_t - C_t. \quad (2.15)$$

Gemäß Annahme 5 kann für die zweite Phase ebenso eine Investitionsquote ermittelt werden. Für sie gilt, dass

$$k_t = k_s \left(1 - \frac{x_t}{T}\right), \quad (2.16)$$

wobei k_t die bereits in Gleichung 2.14 beschriebene Investitionsquote ist, k_s beschreibt das Verhältnis zwischen potentiellm Einkommen und den Opportunitätskosten nach Abschluss der allgemeinen Bildung, x_t sind die Jahre der bisherig generierten Berufserfahrung sowie T die erwartete gesamte produktive Zeit beschreibt. Beim Vergleich der beiden Lebensphasen ergibt sich

2 Theoretische Fundierung

- für die allgemeinbildende Zeit, dass annahmegemäß $k_t = 1$ ist und
- für die Zeit der Akkumulation firmenspezifischen Kapitals gilt, dass nur ein bestimmter Anteil des potentiellen Einkommens investiert wird. Dies bedeutet, dass nach dem Abschluss der allgemeinbildenden Zeit $k_s = \frac{C_s}{E_s} < 1$ ist. Mit zunehmenden Jahren an Berufserfahrung verringert sich die Zeit der möglichen Amortisation der Investition. Demzufolge wird es zunehmend unrentabler, in firmenspezifisches Kapital zu investieren. Ausgedrückt werden kann dies durch den zweiten Term der Gleichung (2.16). Die aktuell geleisteten Berufsjahre ergeben sich als $x_t = t - s$.

Betrachtet man zwei Perioden, so ergibt sich das potentielle Einkommen in Periode t aus dem potentiellen Einkommen der Vorperiode zuzüglich der verzinsten Investition aus dieser Periode. Formal bedeutet dies

$$E_t = E_{t-1} + r_{t-1}C_{t-1}, \quad (2.17)$$

wobei r_{t-1} die Verzinsung der Investition der Vorperiode beschreibt. Ersetzt man C_{t-1} durch den Zusammenhang aus Gleichung (2.14) ergibt sich

$$E_t = (1 + r_{t-1}k_{t-1})E_{t-1}. \quad (2.18)$$

Unter Berücksichtigung der Gleichungen (2.14) und (2.15) ergibt sich

$$Y_t = (1 - k_t)E_t. \quad (2.19)$$

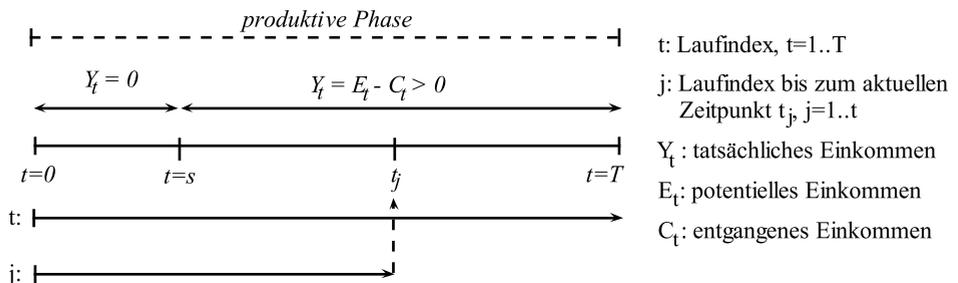


Abbildung 2.6: Indexierung im Modell nach Mincer, eigene Darstellug.

2.2 Einfluss von allgemeinem und firmenspezifischem Humankapital

Um die Zwei-Perioden-Betrachtung zu erweitern, muss ein zusätzlicher Index eingeführt werden. Der bisherige Laufindex t bezieht den gesamten Betrachtungszeitraum der produktiven Phase ein, d. h. von 1 bis T . Der zusätzliche Index j berücksichtigt nun die Perioden bis zum aktuellen Zeitpunkt t . Abbildung 2.6 zeigt diesen Zusammenhang. Die Verallgemeinerung ergibt demnach

$$E_t = \prod_{j=0}^{t-1} (1 + r_j k_j) E_0. \quad (2.20)$$

Unter der Annahme der vollkommenen Reinvestition des potentiellen Einkommens in der allgemeinbildenden Phase sowie der konstanten Verzinsung des zusätzlichen allgemeinen Bildungsjahres, kann die Gleichung umgeformt werden zu

$$E_t = \prod_{j=0}^{s-1} (1 + r_s) \prod_{j=s}^{t-1} (1 + r_j k_j) E_0. \quad (2.21)$$

Die beiden beschriebenen Perioden der produktiven Phase sind nun ersichtlich. Durch Logarithmieren und Anwendung von Summengesetzen erhält man

$$\ln E_t = s \ln(1 + r_s) + \sum_{j=s}^{t-1} \ln(1 + r_j k_j) + \ln E_0. \quad (2.22)$$

Die nächste Umformung unter Anwendung der Logarithmengesetze⁸ ergibt eine näherungsweise Lösung in Form von

$$\ln E_t \approx sr_s + \sum_{j=s}^{t-1} r_j k_j + \ln E_0. \quad (2.23)$$

Unter Einbeziehung der Investitionsquote, formal beschrieben in Gleichung (2.16) sowie der Annahme der stets konstanten Verzinsung eines zusätzlichen Jahres an Berufserfahrung, ergibt sich

$$\ln E_t \approx sr_s + \sum_{j=s}^{t-1} k_s \left(1 - \frac{x_j}{T}\right) r_x + \ln E_0. \quad (2.24)$$

⁸Für kleine Werte von x gilt $\ln(1+x) \sim x$, damit gilt hier $\ln(1+r_s) \sim r_s$ sowie $\ln(1+r_j k_j) \sim r_j k_j$.

2 Theoretische Fundierung

Im Anhang A.1 findet sich eine detaillierte Darstellung der weiteren Umformungen dieser Gleichung. Als Ergebnis der Zwischenrechnungen ergibt sich dann

$$\ln E_{s+x} \approx sr_s + k_s r_x \left(t - 1 - s - \frac{(x_t - 1)x_t}{2T} \right) + \ln E_0.$$

Die potentielle Berufserfahrung im Jahr t ist $x_t = t - s$, damit kann die Gleichung weiter umgeformt werden zu

$$\ln E_{s+x} \approx sr_s + k_s r_x \left(x_t - 1 - \frac{(x_t - 1)x_t}{2T} \right) + \ln E_0.$$

Nach weiteren Umformungen⁹ und Umgruppieren der Gleichung folgt

$$\ln E_{s+x} \approx (\ln E_0 - k_s r_x) + sr_s + \left(k_s r_x \left(1 + \frac{1}{2T} \right) \right) x_t - \left(\frac{k_s r_x}{2T} \right) x_t^2. \quad (2.25)$$

Schlussendlich wird nun Gleichung (2.19) sowie die Gleichung der Investitionsquote berücksichtigt und kann beschrieben werden als

$$\ln Y_{s+x} \approx (\ln E_0 - k_s r_x - k_s) + r_s s + \left(k_s r_x \left(1 + \frac{1}{2T} + \frac{1}{r_x T} \right) \right) x_t - \left(\frac{k_s r_x}{2T} \right) x_t^2. \quad (2.26)$$

Vereinfachend soll nun gelten, dass

$$\ln E_0 - k_s r_x - k_s = \alpha,$$

$$k_s r_x \left(1 + \frac{1}{2T} + \frac{1}{r_x T} \right) = \beta_0,$$

$$\frac{k_s r_x}{2T} = \beta_1.$$

Damit ergibt sich

$$\ln Y_{s+x} \approx \alpha + r_s s + \beta_0 x_t - \beta_1 x_t^2. \quad (2.27)$$

Die Parameter der Gleichung und der Koeffizienten sind definitionsgemäß positiv. Somit kann aus der theoretischen Fundierung ermittelt werden, dass das tatsächliche Ein-

⁹Im ersten Schritt ergibt sich $\ln E_{s+x} \approx sr_s - k_s r_x + k_s r_x \left(x_t + \frac{(1-x_t)x_t}{2T} r_x \right) + \ln E_0$. Durch ausmultiplizieren der Klammern erreicht man $\ln E_{s+x} \approx sr_s - k_s r_x + k_s r_x x_t + \frac{k_s r_x x_t}{2T} - \frac{k_s r_x x_t^2}{2T} + \ln E_0$. Weitere Umgruppierungen führen nun zur Gleichung 2.25.

kommen bestimmt wird durch die Bildungsdauer und die Berufserfahrung. Letztgenannte wirken sowohl als linearer Term einkommenssteigernd als auch im Quadrat einkommensreduzierend auf das Einkommen. Damit gilt, dass das Grenzeinkommen des zusätzlichen Berufsjahres sinkt. Aus dem theoretischen Modell kann die Verzinsung des zusätzlichen Bildungsjahres r_s , die Bildungsrendite, identifiziert werden. Die Verzinsung der Berufserfahrung geht in die Terme β_0 und β_1 mit ein, kann jedoch aufgrund weiterer Einflussgrößen nicht eindeutig identifiziert werden.

2.2.2 Kritik

Ebenso wie das Grundmodell von Mincer beschreibt dieses Modell den Zusammenhang zwischen der generellen Bildung und dem Einkommen. Zusätzlich kann der Einfluss der Berufserfahrung als erklärende Variable mit einbezogen werden. Besonders bemerkenswert ist, dass theoretisch fundiert der auch empirisch vorhandene, positiv abnehmende Grenzertrag der Berufserfahrung dargestellt werden kann.

Jedoch muss die Annahme der ausschließlichen Steigerbarkeit von Humankapital kritisiert werden, da der natürliche Prozess des Vergessens ausgeschlossen wird. Andererseits bezieht sich das Modell auf auszubauendes Humankapital. Es kann davon ausgegangen werden, dass das einkommensgenerierende Wissen in permanenter Benutzung steht und weiterentwickelt wird.

Die Annahme der gleichmäßigen Verzinsung von zusätzlichem allgemeinem und firmenspezifischem Kapital muss nicht gegeben sein. Inwieweit sich das Humankapital aus verschiedenen Bildungsphasen verzinst, kann schwerlich ermittelt werden.

Auch die Einteilung der Lebensphasen mit der deutlichen Trennung beider Phasen erscheint nicht als unbedingt real. Näherungsweise kann diese Annahme jedoch als durchaus sinnvoll erachtet werden, da tendenziell die Reihung in dieser Weise gegeben ist.

Der sich aus Gleichung (2.16) ergebende Zusammenhang für die Investitionsquote stellt lediglich eine Annäherung dar. Es wird dabei davon ausgegangen, dass das Individuum bei Erzielung firmenspezifischen Kapitals potentielles Einkommen nicht erzielen kann. Realistischerweise kann jedoch von einem direkten Lernprozess am Arbeitsplatz ausgegangen werden, wobei sich die Produktivität automatisch durch den täglichen Berufsalltag steigert. Eine Investition in firmenspezifisches Kapital im Sinne von unbezahlten Weiterbildungen außerhalb der Arbeitszeit gibt es zwar, stellt aber eher eine Seltenheit der

beruflichen Praxis dar. Mincer bezieht die Möglichkeit der Arbeitslosigkeit nicht mit ein. Häufig stellt gerade dieses Risiko eine entscheidungsrelevante Größe dar.

Diese Kritikpunkte zeigen deutliche Restriktionen für die Herleitung der Lohngleichung nach Jacob Mincer. Allerdings sind diese teils unrealistischen Annahmen zur Lösbarkeit des Modells zwingend erforderlich. Betrachtet man das Ziel dieses Modells, so soll eine Verbindung zwischen Bildungsjahren und Einkommen geschaffen werden. Diese Ziel erfüllt das Modell und ist demnach zweckmäßig. Mincer liefert damit eine auch empirisch schätzbare Gleichung, die die Zusammenhänge von allgemeinem und firmenspezifischen Humankapital beschreibt. Auch der empirisch häufig nachgewiesene¹⁰ Zusammenhang des positiv abnehmenden Verlaufs der Berufserfahrung auf das Einkommen kann theoretisch hergeleitet werden. Daher ist dieses Modell als theoretische Fundierung und Lieferant für die Ausgangsgleichung der Empirie sinnvoll. Die Heterogenitäten zwischen den Individuen sollen durch die Hinzunahme quantifizierbarer Kontrollvariablen in der Empirie zumindest verringert werden, sodass es möglich ist Schätzungen für die Bildungsrendite zu ermitteln.

2.3 Zusammenfassung

In diesem Kapitel wird die theoretische Fundierung für den Zusammenhang zwischen Humankapital und dem Einkommen gezeigt. Dafür werden verschiedene Modelle von Jacob Mincer diskutiert. Es wird dabei eine diskrete Betrachtung gewählt. Basierend auf dem Vergleich zweier diskontierter Lebensarbeitseinkommen aufgrund verschiedener Bildungsdauern wird das erste Modell hergeleitet. Hierbei wird davon ausgegangen, dass Personen mit geringer Zahl an Bildungsjahren längere Zeit zur Einkommensgenerierung aufwenden können. Mit Blick auf den Einfluss der letzten einkommensgenerierenden Jahre kann die Annahme gleicher individueller Berufsdauern getroffen werden, da durch die Diskontierung der Einfluss der letzten Jahre marginal ist. Unter Verwendung dieser Annahme kann nun als Ergebnis dieses Modells eine einfache Bildungs-Einkommens-Gleichung hergeleitet werden. In einem allgemeineren Modell wird die Lohngleichung um den Einfluss der Berufserfahrung erweitert. Diese dient im weiteren Verlauf dieser Arbeit als Basis für die empirische Analyse.

¹⁰Vgl. u. a. Mincer (1974).

3 Stand der Forschung

Der Einfluss von Bildung ist theoretisch wie empirisch ein vielfach untersuchter Forschungsgegenstand. An dieser Stelle werden einige Studien zur Ermittlung von Bildungsrendite¹¹ in Deutschland gegenübergestellt. Durch die Verwendung verschiedener Datensätze, Methoden und Schätzmodelle können Unterschiede in der geschätzten Bildungsrendite auftreten. Ziel ist es daher, eine Tendenz in der Höhe der Bildungsrenditen zu ermitteln. In den Tabellen 3.1 und 3.2 werden die verwendeten Artikel zusammengefasst.

Beim Vergleich der ausgewählten Studien wird deutlich, dass die Bildungsrendite je nach Methode und Modellvariante zwischen 5,7 % und 16 % schwankt. Die meisten Schätzungen der ermittelten Bildungsrenditen liegen zwischen 6,5 % und 10,2 %. Eine mögliche Ursache dieser Schwankungen ist auf die verwendete Methode zurückzuführen.¹² So wird bei Verwendung der OLS-Methode zumeist eine Rendite bis zu 10 % ermittelt. Im Gegensatz dazu sind die sich ergebenden Werte auf Basis von 2SLS- und 3SLS-Methode um einige Prozentpunkte höher.

Daten. Für die Berechnung der Bildungsrendite wird häufig das Sozio-Ökonomische Panel¹³ des Deutschen Instituts für Wirtschaftsforschung¹⁴ herangezogen. Doch auch der Mikrozensus oder die Daten vom Institut für Arbeitsmarkt- und Berufsforschung¹⁵ liefern eine Grundlage für eine solche Analyse. Lauer und Steiner (2000) und weitere Studien betrachten die Veränderungen der geschätzten Renditen über die Zeit. Dabei lässt sich kein einheitlicher Trend erkennen. Ein Vergleich der Daten aus dem Jahr 1988 der ehemaligen DDR und der Bundesrepublik Deutschland wird von Krueger und Pischke (1995) diskutiert. Dabei verwenden die Autoren die Einkommensstichprobe in Arbeiter- und Angestelltenhaushalten sowie den SOEP-Datensatz. In beiden Ländern ergibt sich eine Rendite

¹¹Die Bildungsrendite ist dabei stets, als eine Verzinsung pro zusätzlichem Bildungsjahr zu verstehen.

¹²Eine genaue Methodenerklärung findet sich in Kapitel 4.

¹³Im Folgenden wird die Kurzform SOEP verwendet.

¹⁴Kurz DIW.

¹⁵Kurz IAB.

3 Stand der Forschung

Studie	Datensatz und Jahr	Bildungsrendite	Bemerkung
Ammermüller und Weber (2005)	SOEP, Mikrozensus; 2002	6,6–8,8 %	Querschnittsanalyse (OLS, Heckman-Methode) für 30–60-Jährige nicht selbständig Beschäftigte in West- und Ostdeutschland
Boockmann und Steiner (2006)	SOEP; 1984–1997	8,4–12,3 %	Panelanalyse (GLS) für deutsche Arbeitnehmer in Westdeutschland im öffentlichen Dienst und der Privatwirtschaft
DiNardo und Pischke (1997)	IAB Betriebspanel; 1979, 1986 und 1992	6,5–7,6 %	Querschnittsanalyse (OLS) der 16–65-Jährigen für 3 verschiedene Jahre in Westdeutschland
Gebel und Pfeiffer (2007)	SOEP; 1984, 1998 und 2002	5,4–7,4 %	Querschnittsanalyse (OLS) der 25–60-Jährigen westdeutschen, nichtselbständig Beschäftigten Vollzeit Arbeitnehmer
Heinicke und Thomsen (2011)	SOEP; 2007–2009 nach Regionen:	8,1–10,6 % 6,8–14,2 %	gepoolter Datensatz der abhängig Beschäftigten mit regionalen Bezügen
Ichino und Winter-Ebmer (1999)	SOEP; 1986	5,7–15,0 %	Querschnittsanalyse (OLS und 2SLS) der männlichen Arbeitnehmer in Westdeutschland
Jochmann und Pohlmeier (2004)	SOEP; 2001	5,5–10,2 %	Querschnittsanalyse (OLS und 2SLS) der 18–47-Jährigen männlichen nicht selbständig Vollzeitbeschäftigten in Westdeutschland
Krueger und Pischke (1995)	Einkommensstichprobe in Arbeiter und Angestelltenhaushalten, SOEP; 1988	7,1–8,9 %	Querschnittsanalyse (OLS) der nicht selbständig Vollzeitbeschäftigten in der BRD und der ehemaligen DDR
Lauer und Steiner (2000)	SOEP; 1984–1997	6,0–10,6 %	Längs- und Querschnittsanalyse (OLS, 2SLS sowie die Heckman-Methode) der 30–60-Jährigen Arbeitnehmer in Westdeutschland

Tabelle 3.1: Bildungsrenditen in Deutschland Teil 1, Berechnung basierend auf der Lohngleichung nach J. Mincer.

Studie	Datensatz und Jahr	Bildungsrendite	Bemerkung
Maier et al. (2004a)	IAB Strukturhebung; 1999	8,7 %	Querschnittsanalyse (2SLS) der nicht selbständig vollzeitbeschäftigten Männer in Westdeutschland
Reilich (2006)	SOEP; 2004	7,9–13,6 %	Querschnittsanalyse (OLS, 2SLS, 3SLS und Heckman-Methode) der 30–55-Jährigen nicht selbständig Beschäftigten in Deutschland

Tabelle 3.2: Bildungsrenditen in Deutschland Teil 2, Berechnung basierend auf der Lohngleichung nach J. Mincer.

in Höhe von 8 %. Ammermüller und Weber (2005) vergleichen direkt zwei Datensätze miteinander. Dabei werden die Ergebnisse des Jahres 2002 auf Basis der SOEP-Daten mit jenen des Mikrozensus aus dem Jahr 2000 jenen gegenübergestellt. Beide Renditen werden mit ähnlichen Variablen und Eingrenzungen geschätzt. So liegt die Rendite, berechnet aus den Daten vom Mikrozensus zwischen 6,6 % und 7,6 %, die über die SOEP-Daten ermittelten Renditen liegen zwischen 6,9 % und 7,7 %.

Methoden. Betrachtet man zunächst ausschließlich jene Autoren, die das OLS-Verfahren angewandt haben, ergibt sich insgesamt eine Rendite zwischen 5,4 % und 10,3 %. Die Ergebnisse der Querschnittsanalyse beziehen sich auf ausgewählte Jahre zwischen 1979¹⁶ und 2009¹⁷. Die meisten Studien schätzen unter Anwendung des OLS-Verfahrens eine Rendite zwischen 6,5 % und 8,0 %. Ein besonderer Fokus wird häufig auf die sich ergebenden Unterschiede der Bildungsrendite aufgrund des Geschlechts, des Alters oder der Herkunft (Ost- oder West-Region) des Individuums gelegt.

Tendenziell höhere Renditen liefert die 2SLS-Methode¹⁸. Die geringste Rendite messen mit dieser Methode Ichino und Winter-Ebmer (1999) mit 4,9 % im Jahr 1986. Die höchste Rendite ergibt sich mit 16,0 % bei Lauer und Steiner (2000). In beiden Artikeln werden die Daten des SOEP zugrunde gelegt. Auffällig in der erstgenannten Studie ist jedoch, dass die Modellspezifikation der Bildungsgleichung einen deutlichen Einfluss auf die Ergebnisse hat. So messen Ichino und Winter-Ebmer (1999) eine Rendite von 4,9 % bzw., bei geringer

¹⁶Vgl. DiNardo und Pischke (1997).

¹⁷Vgl. Heinicke und Thomsen (2011).

¹⁸Mehr Details zu dieser Methodik siehe Abschnitt 5.2.1.

3 *Stand der Forschung*

Abwandlung der ersten Stufe des Regressionsmodells, eine Rendite von 15 %. Diese große Variabilität zeigt, dass ein besonderes Augenmerk auf die Auswahl geeigneter Instrumente gelegt werden muss. Beim Vergleich zwischen dem 2SLS- und 3SLS-Verfahrens ergeben sich keine wesentlichen Änderungen.¹⁹

Im Gegensatz hierzu liefert die Heckman-Methode geringere Renditen. So ermitteln Ammermüller und Weber (2005) mit 6,6 % die geringsten Renditen mit diesem Verfahren. Lauer und Steiner (2000) errechnen Renditen im Intervall von 7,4–8,9 %. Höher dagegen sind die Renditen von Reilich (2006) die in einer Spanne zwischen 9,4 % und 10,8 % liegen. Im Vergleich zu den Ergebnissen, die sich aus dem 2SLS-Verfahren ergeben, sind die Renditen, ermittelt durch die Heckman-Methode, jedoch deutlich geringer.

Eine weitere Möglichkeit der Berechnung von Bildungsrenditen liefert die Längsschnittanalyse. Hierbei werden von Boockmann und Steiner (2006) Renditen für Männer in Höhe von 8,4 %, für Frauen in Höhe von 12,3 % ermittelt. Der betrachtete Zeitraum liegt zwischen 1984 und 1997. Für den gleichen Zeitraum ermitteln Lauer und Steiner (2000) für Männer eine Rendite von 8,3 % und für Frauen von 10,5 % pro zusätzlichem Bildungsjahr. Heinicke und Thomsen (2011) wählen einen deutlich kürzeren Zeitraum. Sie ermitteln für die Jahre 2007 bis 2009 eine Rendite für Männer in Westdeutschland in Höhe von 9,4 %, in Ostdeutschland in Höhe von 10,6 %. Die Rendite für Frauen ist, anders als bei den letztgenannten Studien, in beiden Gebieten Deutschlands geringer als für Männer. So wird eine Frau in Westdeutschland eine Rendite in Höhe von 8,1 % und im Osten Deutschlands in Höhe von 8,4 % erzielen. Interessant ist bei dieser Studie die Konzentration auf die regionalen Unterschiede in den Bildungsrenditen. Berücksichtigt werden jedoch nur einzelne ausgewählte²⁰ Bundesländer. Die Analyse wird wiederum nach dem Geschlecht separiert vorgenommen. Die geringste Rendite wird in Mecklenburg-Vorpommern für Frauen berechnet. Sie erhalten lediglich eine Verzinsung pro zusätzlichem Bildungsjahr von 6,8 %. Die höchste Rendite ist im gleichen Bundesland berechnet worden. Männer in Mecklenburg-Vorpommern haben eine Rendite von durchschnittlich 14,2 %. Eine gleiche Rendite für Männer und Frauen konnte für Schleswig-Holstein berechnet werden. Hier kann unabhängig vom Geschlecht von einer Verzinsung in Höhe von 12,1 % ausgegangen werden.

Ausgehend von den dargestellten Studien wird deutlich, dass die Renditen in der Höhe zumeist ähnlich sind, sich jedoch aufgrund bestimmter Spezifikationen unterscheiden kön-

¹⁹Vgl. Reilich (2006).

²⁰Ausgewählt wurden dabei Schleswig-Holstein, Hessen, Bayern, Mecklenburg-Vorpommern, Sachsen und Sachsen-Anhalt.

nen. Offensichtlich spielt der aktuelle Wohnort eine entscheidende Rolle für die Bildungsrendite. Diese Unterscheidung wurde in einigen berücksichtigt, indem Renditen für Ost- und Westdeutschland getrennt geschätzt wurden. Heinicke und Thomsen (2011) erweitern diesen Gedanken und untergliedern weiter in ausgewählte Unterregionen. Ein Ziel dieser Arbeit ist es, diese Untergliederung zu erweitern und eine Schätzung der Bildungsrenditen für jedes Bundesland implizit vorzunehmen. Hierfür gilt es, geeignete Methoden zu identifizieren und die Ergebnisse in sinnvoller Weise zu vergleichen.

Kritik. Dieser Überblick basiert auf einer Auswahl bestehender Literatur zur empirischen Schätzung von Bildungsrenditen in Deutschland. Dabei ist eine Auswahl mit Blick auf die Methode und die Region getroffen worden. Aufgrund der Vielzahl der vorhandenen Literatur kann diese nicht vollumfänglich wiedergegeben werden. Demnach stellt dieser Abschnitt einen Abriss über einen Teil der vorhandenen Literatur dar. Ziel bei der Auswahl der Studien war es, einen Überblick über die Höhen der geschätzten Bildungsrenditen, die verwendeten Methoden sowie den aktuellen Stand der Diskussion der regionalen und gesamtdeutschen Analyse zu liefern.

Zusammenfassung. *Viele Studien liefern Ergebnisse zur Schätzung der Bildungsrenditen für Deutschland. Die Ergebnisse ähneln sich. Unabhängig von den Methoden oder den Daten werden von den meisten Studien Bildungsrenditen im Bereich zwischen 6,5 % und 10,2 % ermittelt. Beim Vergleich der Ergebnisse zeigt sich die größte Variabilität der Ergebnisse bei Verwendung der 2SLS-Methodik. Der Vergleich zwischen Längs- und Querschnittsstudien zeigt keine Unterschiede in der Höhe der Bildungsrendite. Viele Studien nehmen Separierungen bestimmter Charakteristika vor. Ein Ziel dieser Arbeit ist die Erweiterung im Forschungsgebiet der regionalen Bildungsrenditen.*

4 Daten und Datenbeschreibung

In diesem Kapitel wird der Datensatz des DIW dargestellt sowie die wesentlichen Variablen für die in Kapitel 5 dargestellte Analyse vorgestellt. Dabei wird zunächst der Datensatz allgemein beschrieben sowie die mit Blick auf die Forschungsfrage notwendigen durchgeführten Reduktionen werden diskutiert. Im Weiteren erfolgt eine Darstellung der wesentlichen Variablen mit einer bivariaten Analyse den im Folgenden angenommenen Zusammenhänge.

4.1 Datensatzeingrenzungen

Dieser Abschnitt stellt zunächst den allgemeinen Datensatz sowie die notwendigen Reduktionen aus diesem vor. Es folgt eine ausführliche Repräsentativitätsprüfung um die mögliche Übertragbarkeit der Ergebnisse auf die Grundgesamtheit darzustellen.

4.1.1 Allgemeiner Datensatz

*„Das SOEP ist eine repräsentative Wiederholungsbefragung privater Haushalte in Deutschland, die im jährlichen Rhythmus seit 1984 bei denselben Personen und Familien in der Bundesrepublik durchgeführt wird.“*²¹ Ähnlich wie das US-amerikanische Panel Study of Income Dynamics (PSID) ist diese Befragung als Längsschnitt angelegt um Mikrodaten mit Veränderungen u. a. in der Einkommensstruktur oder der Haushaltszusammensetzung über die Jahre beobachten zu können. Anders als beim PSID werden bei der Datenerhebung des Sozioökonomischen Panels (SOEP) alle²² der zufällig²³ ausgewählten Haushaltsmitglieder befragt. Die Befragung erfolgt über ein persönliches Interview.

²¹DIW, 2011. Homepage: www.diw.de/deutsch/soep/26628.html. Abgerufen am 25.07.2011.

²²Das PSID befragt lediglich eine Person pro Haushalt.

²³Siehe hierzu DIW, SOEP-Dokumentation, 2011.

Wagner et al. (2007) beschreiben die über die Jahre schrittweisen Anpassungen des Datensatzes. Dabei werden Modifizierungen in der Art der Fragen und auch im Stichprobenumfang vorgenommen. Ziel bei der Auswahl der Befragten ist stets die Sicherstellung einer für Deutschland repräsentativen Stichprobe.

In dieser Arbeit wird eine Querschnittsanalyse mit den Daten aus dem Erhebungsjahr 2008 vorgenommen. Hierbei umfasste die Stichprobe fast 11.000 Haushalte mit insgesamt mehr als 20.000 Personen. Befragt werden die Mitglieder privater Haushalte, die das 16. Lebensjahr erreicht haben.

4.1.2 Datensatz der Erwerbstätigen

Zur Berechnung der Bildungsrendite²⁴ wird der Ansatz von Mincer (1974) angewendet. Dabei sind ausschließlich nichtselbstständig Beschäftigte in die Untersuchung einbezogen, die zumindest in Teilzeit²⁵ beschäftigt sind.

Merkmal	Stichprobenumfang
aus Grunddatei, nur Voll- und Teilerwerbstätige, die Angaben zur Bildungsdauer gemacht haben und nicht selbstständig beschäftigt, bzw. noch im Ausbildungsprozess sind	8.719
Altersbeschränkung auf das Intervall 30–60 Jahre	7.119
ohne fehlende Antworten	4.693

Tabelle 4.1: Reduktion aus der Grunddatei; Quelle: DIW, eigene Berechnungen. Jahr: 2008.

Einige Datensatzreduzierungen²⁶ mussten aufgrund von fehlenden Antworten vorgenommen werden. Eine weitere – in der Literatur übliche – Reduktion ist durch eine Altersbeschränkung gegeben. Die hier durchgeführte Analyse bezieht sich auf Individuen zwischen dem 30. und 60. Lebensjahr. Grund dieser Einschränkung ist, dass für eine hohe Qualifikation ein langer Bildungsweg vonnöten ist und damit viele unter 30-Jährige, die ein akademisches Niveau anstreben, zumeist noch im Ausbildungsprozess sind. Die obere Grenze ist notwendig, da viele Erwerbstätige nicht bis zum Renteneintrittsalter tätig

²⁴Die Bildungsrendite ist dabei stets, als eine Verzinsung pro zusätzlichem Bildungsjahr zu verstehen.

²⁵Diese Einschränkung gilt nicht vollumfänglich für die vorliegende Arbeit. In Abschnitt 5.4.2 soll auf mögliche Selbstselektionsproblematiken und die Anwendbarkeit der Berechnung der Bildungsrendite auch für einen Datensatz mit Arbeitslosen vorgenommen werden. Demzufolge wird es in diesem Abschnitt eine Erweiterung des Datensatzes geben.

²⁶Die Daten wurden mit den Programmen StataIC 12, SPSS 19 und R (2.11.1) aufbereitet und ausgewertet.

sind. Die Auswirkung der Reduktionen sind in Tabelle 4.1 mit den zugehörigen Stichprobenumfängen dargestellt. Berücksichtigt man alle genannten Einschränkungen, reduziert sich der Datensatz auf 4.693 Individuen. Insgesamt werden unter Einbezug der Alters- und Berufsbeschränkung 0,023 % der Grundgesamtheit²⁷ in der Analyse betrachtet.

4.1.3 Repräsentativitätsprüfung des Datensatzes

Um die Ergebnisse aus der Analyse auf die Grundgesamtheit übertragen zu können, muss der Datensatz diese repräsentieren. Dafür muss *„die Auswahl einer Teilgesamtheit [...] so [...] vorgenommen werden, dass aus dem Ergebnis der Teilerhebung möglichst exakt und sicher auf die Verhältnisse der Gesamtmasse geschlossen werden kann. Dies ist dann der Fall, wenn die Teilerhebung in der Verteilung aller interessierenden Merkmale der Gesamtmasse entspricht, d. h. ein zwar verkleinertes, aber sonst wirklichkeitsgetreues Abbild der Gesamtheit darstellt.“*²⁸ Zur Prüfung der Verallgemeinerungswürdigkeit der Ergebnisse werden zwei Sachverhalte geprüft:

Zum einen muss die Auswahl der Befragten zufällig stattfinden. Nur dann kann von unabhängig und identisch verteilten Zufallsvariablen ausgegangen werden. Da die hier verwendeten Daten vom DIW erhoben werden, kann eine genaue Prüfung der Datenerhebung nicht erfolgen. Aus der Dokumentation²⁹ des DIW geht jedoch die hohe Sensibilität bei der Auswahl hervor. Eine Prüfung des allgemeinen SOEP-Datensatzes wird von Schröppler (2011) vorgenommen. Er untersucht den Datensatz unter Zuhilfenahme des Gesetzes von Benford³⁰. Hierbei wird eine allgemeine Verteilung der Anfangsziffern von Zahlen angenommen. Benford weist nach, dass in vielen empirischen Daten kleine Anfangsziffern überproportional vertreten sind. Schröppler (2011) zeigt nun, dass dieses Gesetz auf monetäre Variablen im SOEP-Datensatz anwendbar ist. So vergleicht er zum Beispiel das Bruttoeinkommen für verschiedene Jahre. Dabei untersucht er die zugehörige Verteilung der Erstziffern. In den Jahren der Wiedervereinigung sind die niedrigen Erstziffern noch deutlich mehr vertreten, als Benford es ursprünglich ermittelt hat. Doch im Zuge der Angleichung wird für die nachfolgenden Jahre eine Annäherung an die relativen Häufigkeiten von Benford gezeigt. Demzufolge kann für die monetären Werte davon ausgegangen wer-

²⁷Die Daten für die Grundgesamtheit lieferte die Bundesagentur für Arbeit mit der nötigen Alterseingrenzung. Aus: www.arbeitsagentur.de/Statistiken. Abgerufen am 25.07.2011.

²⁸Von der Lippe und Kladroba (2002), S. 3.

²⁹Siehe hierzu DIW, SOEP-Dokumentation, 2011.

³⁰Näheres siehe Benford (1938).

den, dass der Grunddatensatz aus dem Jahr 2008 dem Repräsentativitätsanspruch gerecht wird.

Zum anderen muss geprüft werden, ob der reduzierte Datensatz noch immer dem Repräsentativitätsanspruch für die zu untersuchende Gruppe genügt. Für diese Prüfung werden nun einige deskriptive Statistiken des statistischen Bundesamtes³¹ mit den Ergebnissen der hier verwendeten Stichprobe verglichen. Als Vergleichsstichprobe werden die Daten des Mikrozensus des Jahres 2008³² herangezogen. Bei dieser Stichprobe handelt es sich um eine Erhebung im Rahmen der amtlichen Statistik. Dabei wird 1 % der deutschen Bevölkerung durch das Klumpenverfahren repräsentativ ausgewählt und befragt. Aufgrund der gegebenen gesetzlichen Regelungen der amtlichen Statistik sind die ausgewählten Individuen zur Auskunft³³ verpflichtet. Daher kann für diese Stichprobe eine Ausfallquote von weniger als 5 % festgestellt werden. Bei dieser hohen Rücklaufquote kann davon ausgegangen werden, dass in der Erhebung nahezu keine Verzerrungen aufgrund selbstselektierenden Verhaltens zu finden sind. Durch die Auskunftspflicht kann, bei gegebener Zufälligkeit, von einer Übertragbarkeit der Mikrozensus-Daten auf die Grundgesamtheit ausgegangen werden. Ähnlich wie bei dem vorliegenden SOEP-Datensatz muss auf die Zufälligkeit der Auswahl der Beobachtungen durch das statistische Bundesamt bzw. der statistischen Landesämter vertraut werden. Generell kann der Vergleich mit den Daten des statistischen Bundesamtes aufgrund der Auskunftspflicht als geeignet angesehen werden. Eine perfekte Vergleichbarkeit kann häufig jedoch nicht hergestellt werden. Dies wird zum Ende dieses Abschnitts ausführlich diskutiert.

Vergleich der Stichproben. Ziel ist es, einige sinnvolle³⁴ deskriptive Statistiken zwischen dem SOEP-Datensatz und dem Mikrozensus zu vergleichen. Dabei wird speziell auf Häufigkeitsverteilungen und Mittelwertvergleiche eingegangen. Zur Durchführung muss beachtet werden, dass der verwendete SOEP-Datensatz des DIW auf Arbeitnehmer zwischen dem 30. und 60. Lebensjahr reduziert wurde. Weiterhin beziehen sich alle Daten auf das Erhebungsjahr 2008. Zur sinnvollen Vergleichbarkeit muss demzufolge die Vergleichs-

³¹ Abgekürzt in Verweisen mit destatis.

³² Siehe hierzu <http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Presse/abisz/Mikrozensus,templateId=renderPrint.psml>. Abgerufen am 12.12.2011.

³³ Einige Fragen sind von der Auskunftspflicht ausgenommen. Näheres dazu ebd. S. 3.

³⁴ Sinnvoll bedeutet hier, dass die Daten beidseitig verfügbar und auch in großem Maß den gleichen Abgrenzungen unterliegen.

stichprobe möglichst genau³⁵ auf diese Beobachtungen reduziert werden. Bei Altersgruppenunterscheidungen ist es in den Daten des statistischen Bundesamts fast nie möglich, die Personen im 60. Lebensjahr mit einzubeziehen. Einzige Ausnahme ist hierbei die Berechnung des Altersdurchschnitts. Hierfür kann vom statistischen Bundesamt auch das 60. Lebensjahr mit einbezogen werden. Es ist jedoch für die anderen Merkmale davon auszugehen, dass der Einfluss des letzten betrachteten Lebensjahres eher marginaler Natur ist und demzufolge eine annähernde Vergleichbarkeit hergestellt werden kann. Die genaue Beschreibung der verwendeten Variablen des SOEP-Datensatzes mit zugehörigen deskriptiven Statistiken findet sich in Abschnitt 4.2.

Monatliches Einkommen. Das statistische Bundesamt liefert eine Aufschlüsselung der monatlichen Bruttoeinkommen nach diversen Kriterien. Leider ist eine Aufteilung nach Altersclustern für diese Variable nicht zu erhalten. Dafür wird das mittlere monatliche Bruttoeinkommen je Wirtschaftszweig zur Verfügung gestellt. Im Mittel verdient ein Erwerbstätiger hierbei 2.847 € brutto im Monat. Im Vergleichsdatsatz verdient eine Arbeitnehmer zwischen 30 und 60 Jahren im Mittel 2.728 €. Diese sehr geringe Abweichung von 4 % deutet auf ähnliche Einkommen hin.

Bildungsniveau. In der empirischen Analyse dieser Arbeit wird die Bildungsdauer verwendet. Diese, in Jahren gemessene Variable wird vom DIW auf Basis der Abschlüsse generiert. Hierfür gibt es im Mikrozensus keine adäquate Vergleichsgröße. Erhältlich in beiden Datensätzen dagegen ist der Abschluss der schulischen Ausbildung. Es ist davon auszugehen, dass die Bildungsdauer und das Bildungsniveau stark korrelieren. Daher eignet sich das Bildungslevel gut als Vergleichsgröße. Eine Alterseinschränkung auf das Intervall von 30 bis 59 Jahren ist auch für den Mikrozensus erhältlich.

Abbildung 4.1 zeigt die beiden sich ergebenden Verteilungen für die Schulabschlüsse der Erwerbstätigen in Deutschland. Die Anteile der Erwerbstätigen mit Abitur sind fast gleich groß. Der Anteil der Personen ohne Schulabschluss spielt nur eine verschwindend geringe Rolle. Im Vergleich ist der Anteil der Personen mit einem Hauptschulabschluss im Datensatz des statistischen Bundesamtes ca. 5 % höher. Die Personen mit Realschulabschluss sind dagegen im SOEP-Datensatz leicht überrepräsentiert. Abgesehen von diesen kleineren Abweichungen sind die Verteilungen jedoch ähnlich.

³⁵Dies ist leider aus Verfügbarkeitsgründen nicht vollumfänglich gegeben. Es wird jedoch an jeder Stelle gezielt darauf verwiesen.

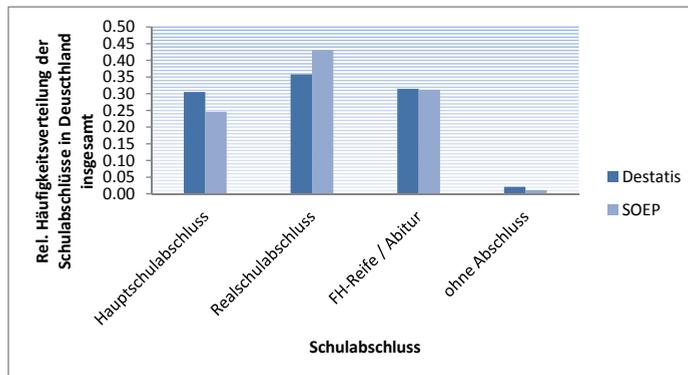


Abbildung 4.1: Verteilung der Schulabschlüsse in Deutschland im Vergleich; Quelle: Statistisches Bundesamt, DIW, eigene Berechnungen.

Zusätzlich werden diese Verteilungen auch für die einzelnen Bundesländer aufgeschlüsselt. Die Grafiken pro Bundesland sind im Anhang A.2.1 dargestellt. Sie zeigen eine deutliche Heterogenität der Verteilung der Schulabschlüsse zwischen den Bundesländern. Der Anteil der Personen mit einem Hauptschulabschluss ist in den neuen Bundesländern deutlich geringer als in den alten Bundesländern. Der Anteil der Erwerbstätigen mit Abitur variiert stark. So haben die Stadtstaaten mit jeweils über 40 % einen deutlich größeren Anteil an Personen mit Universitätszugangsberechtigung. Deutlich niedriger dagegen sind die Anteile in den Flächenländern der ehemaligen DDR. Diese Heterogenität zeigen beide Stichproben gleichermaßen. Die Verteilungen im Vergleich der Prozentwerte sind nicht immer identisch, weisen doch aber im Groben ähnliche Werte auf.

Durchschnittsalter. Beim Vergleich beider Stichproben hinsichtlich des Alters kann nur eine sehr geringe Abweichung festgestellt werden. Das mittlere Alter bei den 30 bis 60-Jährigen³⁶ in der Befragung des statistischen Bundesamtes liegt bei 45,7 Jahren. Das SOEP zeigt ein mittleres Alter von 45,4 Jahren.

³⁶Diesmal konnten auch bei destatis die 60-Jährigen in der Betrachtung berücksichtigt werden.

Betrachtet man das Durchschnittsalter separiert nach dem Geschlecht, so ergibt sich ein ähnliches Bild. Männer sind im Durchschnitt 45,6 Jahre im Datensatz des statistischen Bundesamtes und 45,2 Jahre in der verwendeten Stichprobe. Die Durchschnittsalter der Frauen unterscheiden sich lediglich um ein Vierteljahr, wobei auch in diesem Fall die befragten Frauen in der Stichprobe des statistischen Bundesamtes älter sind.

Geschlecht. Viele Unterschiede ergeben sich aufgrund unterschiedlicher geschlechtsspezifischer Zusammensetzung. Abbildung 4.2 zeigt den Anteil der männlichen Erwerbstätigen pro Altersgruppe. Generell kann festgestellt werden, dass in jeder Altersgruppe und in beiden Stichproben mindestens 50 % der Personen männlich sind. Die Unterschiede zwischen den beiden Stichproben sind gering, sodass davon ausgegangen werden kann, dass die Geschlechterverteilung näherungsweise auf die Grundgesamtheit³⁷ übertragbar ist.

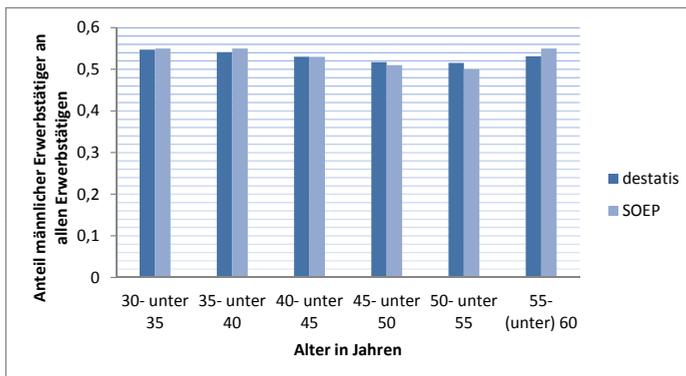


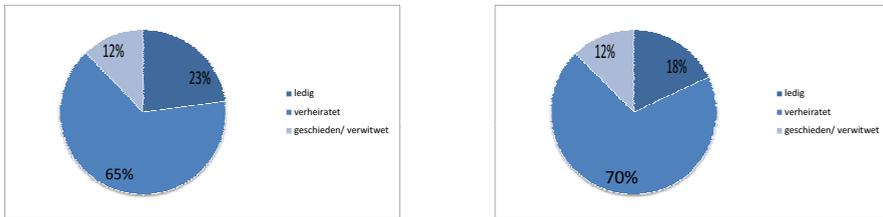
Abbildung 4.2: Anteil männlicher Erwerbstätiger nach Altersgruppen im Vergleich; Quelle: statistisches Bundesamt, DIW, eigene Berechnungen.

Familienstand. Der Familienstand liefert, als eine leicht erhebbare und gut vergleichbare Variable, ein weiteres nützliches Vergleichsinstrument. Abbildung 4.3 zeigt zunächst

³⁷Unter der bereits gesetzten Annahme, dass der Mikrozensus die Grundgesamtheit approximativ widerspiegelt.

4 Daten und Datenbeschreibung

eine Darstellung der relativen Häufigkeitsverteilung des Familienstandes der 30 bis (unter) 60-Jährigen. In der Auswahl der Mikrozensusdaten steht entweder nur die Aufschlüsselung des Familienstandes der Erwerbspersonen in der Abgrenzung sinnvoller Altersgruppen zur Verfügung, oder aber es gibt eine Aufschlüsselung der Erwerbstätigen ohne Altersgruppen. Da die Darstellung des Familienstands ohne Berücksichtigung der Altersgruppen wenig sinnvoll ist, wird hier auf die Erwerbspersonen zurückgegriffen. Hierbei ist zu berücksichtigen, dass diese Basis sich von der des SOEP-Datensatzes um die Erwerbslosen unterscheidet. Beim Vergleich der Anteile ergibt sich, dass die geschiedenen bzw. verwitweten Personen mit 12 % in beiden Stichproben gleich häufig auftreten. Ein Unterschied ergibt sich bei ledigen bzw. verheirateten Individuen.



(a) Anteile der Erwerbspersonen; Quelle: statistisches Bundesamt, eigene Berechnungen. (b) Anteile der Erwerbstätigen; Quelle: DIW, eigene Berechnungen.

Abbildung 4.3: Häufigkeitsverteilungen des Familienstands der 30–60-Jährigen im Vergleich.

Der Anteil der Verheirateten ist im SOEP-Datensatz um 5 % höher als in den Daten des Mikrozensus. Dieser Anteil lässt sich dagegen im letztgenannten Datensatz den ledigen Personen zuordnen.

Erwerbstätige pro Bundesland. Um die Repräsentativität für die Bundesländer zu prüfen, wird die relative Häufigkeitsverteilung der Erwerbstätigen³⁸ auf die Bundesländer dargestellt und verglichen. Generell ist festzustellen, dass Baden-Württemberg und Bayern in dem verwendeten Datensatz leicht unterrepräsentiert sind. Dagegen sind Sachsen, Sachsen-Anhalt und Thüringen eher überrepräsentiert. Allgemein kann jedoch durch den Vergleich eine ähnliche Verteilung auf die Bundesländer identifiziert werden. Die Abweichungen sind gering.

³⁸Ohne Berücksichtigung der Altersgruppen im Mikrozensus.

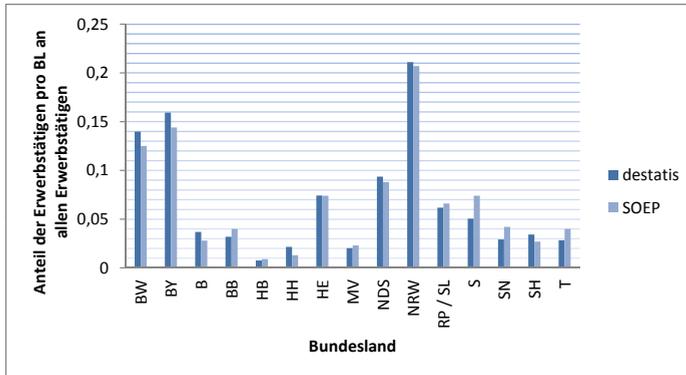


Abbildung 4.4: Verteilung der Erwerbstätigen auf die Bundesländer im Vergleich; Quelle: statistisches Bundesamt, DIW, eigene Berechnungen.

Weitere Statistiken. Vergleicht man den Anteil der Teilerwerbstätigen der interessierenden Altersgruppe an allen Erwerbstätigen, ist ersichtlich, dass im Durchschnitt 22 % der Erwerbstätigen, die durch das statistische Bundesamt befragt wurden, nicht in Vollzeit arbeiten. In den SOEP-Daten sind es 23 %.

Die mittlere wöchentliche Arbeitszeit der abhängig Beschäftigten liegt in beiden Stichproben bei 34,5 h die Woche. Dabei ist bei den Daten des statistischen Bundesamtes keine altersgruppenspezifische Eingrenzung vorhanden.

Kritik und Zusammenfassung der Ergebnisse. Die Repräsentativitätsprüfung besteht aus zwei Teilen: a) die Zufälligkeit der Datenerhebung und b) die Verallgemeinerungsmöglichkeit nach der Reduktion des Datensatzes. Bezüglich der ersten Fragestellung muss auf die Zufälligkeit der Erhebung vertraut werden, da diese extern durchgeführt wird. Indizien, wie die Untersuchung von Schräpler (2011), deuten grundsätzlich auf eine zufällige Beobachtungsauswahl hin. Die zweite Frage nach der Repräsentativität für den reduzierten Datensatz wird über den Vergleich deskriptiver Statistiken geprüft. Dieser Vergleich zweier Stichproben gestaltet sich in der Umsetzung zumeist als schwierig, da häufig die vorgenommenen Eingrenzungen in einem Datensatz mit jenen des anderen nicht über-

einstimmen, bzw. eine genaue Gleichheit aufgrund nicht gegebener Verfügbarkeit zumeist schlecht herzustellen ist. Für diese Prüfung werden sinnvolle Vergleichsvariablen gewählt. Dabei ist die Auswahl der Variablen mit Blick auf die Verfügbarkeit und die Vergleichbarkeit zu treffen. Diese Auswahl stellt nicht sicher, dass es nicht weitere Variablen im Datensatz gibt, die sich grundsätzlich unterscheiden. Reduzieren lässt sich dieses Risiko durch eine recht breitgefächerte Auswahl. Einige verwendete Variablen sind nur teilweise vergleichbar, zum Beispiel im Hinblick auf die verschiedenen Alterscluster. So muss zum Beispiel das Bruttoeinkommen im Vergleichsdatsatz für alle Erwerbstätigen, also ohne Alterseingrenzung, verwendet werden. Der Vergleich des Familienstands ergibt sich im SOEP aus den Erwerbstätigen, im Mikrozensus aus den Erwerbspersonen. Einige Unterschiede in der Ausprägung der Vergleichsgrößen sind nicht von der Hand zu weisen. Es wäre unwahrscheinlich, wenn identische Ergebnisse aus zwei Stichproben erzielt werden könnten. Die Tendenz der Verteilungen ist jedoch bei allen Vergleichsgrößen ähnlich. Die Durchführung von Signifikanztests ist leider nicht möglich, da das statistische Bundesamt in den meisten Fällen die ausgewerteten Daten liefert und keine Rohdaten verfügbar sind. Wenn man davon ausgeht, dass der Datensatz des statistischen Bundesamtes, aufgrund des verpflichtenden Antwortverhaltens, näherungsweise repräsentativ ist, kann diese Ähnlichkeit als Indiz für eine gegebene Repräsentativität verstanden werden. Ein sicherer Wert für die Grundgesamtheit kann nicht durch die Analyse einer Teilerhebung ermittelt werden. Demzufolge ist dies nicht der Anspruch dieser Arbeit. Es sollen jedoch alle Aussagen und Ergebnisse, die im Rahmen der statistischen Möglichkeiten ermittelt werden, als Näherungswerte für die Grundgesamtheit verstanden werden.

Die im Rahmen dieser Arbeit getroffenen Aussagen können sich jeweils nur auf die Individuen im betrachteten Personenkreis beziehen. So wird innerhalb der Analyse die Gruppe der nichtselbstständigen Erwerbstätigen zwischen dem 30. und 60. Lebensjahr betrachtet. Aufgrund der Reduktion auf diese Teilgesamtheit wird die Gruppe derer, die keiner³⁹ Arbeit nachgehen oder selbstständig beschäftigt sind, ausgeschlossen.

³⁹Ausnahme bildet hier Abschnitt 5.4.2. Hier wird der Datensatz um Arbeitslose erweitert und eine mögliche Erweiterung der Ergebnisse geprüft.

4.2 Beschreibung der Variablen und deskriptive Analyse

Dargestellt werden Variablen⁴⁰ und die zugehörigen deskriptiven Statistiken basierend auf dem in Abschnitt 4.1 beschriebenen Datensatz. Die Beurteilung der Signifikanz erfolgt stets mit einer Irrtumswahrscheinlichkeit von 5 %. Der Vollständigkeit halber werden weitere Signifikanzniveaus⁴¹ (1 % und 10 %) mit angegeben. Ziel ist es, im weiteren Verlauf diese Variablen als Regressor oder Regressand in Schätzmodelle einzubauen. Hauptgegenstand der Arbeit ist die Ermittlung einer Bildungsrendite über die Schätzung einer Lohngleichung. Um eventuelle Verzerrungen⁴² zu vermeiden, wird zusätzlich eine Schätzung der Bildungsdauer vorgenommen. Je nachdem in welche Regression die jeweils betrachtete Variable einbezogen ist, wird ein Mittelwerttest zur jeweils abhängigen Variable in diesem Abschnitt gezeigt. Im letzten Schritt werden drei den Datensatz separierende Variablen diskutiert. Für diese Untergruppen wird im Weiteren die Bildungsrendite getrennt ermittelt.

4.2.1 Variablen zur Schätzung der Lohngleichung

Dieser Abschnitt beschreibt die Variablen zur Schätzung der Lohngleichung. Zunächst wird dabei auf das Bruttoerwerbseinkommen und dessen Verwendung eingegangen, um im Weiteren die Einflussgrößen für diesen zu diskutieren. Dabei wird die Motivation zu Verwendung der einzelnen Variablen gezeigt sowie die Zusammenhänge in einem bivariaten Mittelwertvergleich dargestellt.

Aktuelles Bruttoerwerbseinkommen. Das Bruttoerwerbseinkommen ist die monatliche individuelle Entlohnung vor Steuern für den Arbeitnehmer in der produktiven⁴³ Lebensphase. Zur besseren Vergleichbarkeit wird diese Variable umgewandelt in den Bruttoerwerbslohn pro Stunde. Die Berechnung erfolgt durch:

$$\text{Bruttostundenlohn} = \frac{\text{aktuelles Bruttoerwerbseinkommen}}{\text{Anzahl der Wochen pro Monat} \cdot \text{vereinbarte Wochenarbeitszeit}}$$

⁴⁰Dieser Abschnitt basiert teilweise auf Reilich (2012).

⁴¹Auf diese zusätzliche Angabe wird in seltenen Fällen aus Gründen der Übersichtlichkeit verzichtet.

⁴²Näheres siehe Abschnitt 5.2.1.

⁴³Vgl. Abschnitt 2.1.

Ausgehend von einer durchschnittlichen Wochenanzahl von 4,35 Wochen pro Monat wird das Einkommen korrigiert. Die Wochenarbeitszeit wird durch das DIW als separate Frage erhoben. Zur Auswahl stehen dabei die tatsächliche und die vereinbarte Wochenarbeitszeit. In dieser Arbeit wird die vertraglich vereinbarte Arbeitszeit berücksichtigt. Diese ist geeigneter, da zum einen davon auszugehen ist, dass die Entscheidung für oder gegen ein Jobangebot auf Basis der vereinbarten Arbeitszeit getroffen wird und zum anderen die tatsächliche Arbeitszeit stark volatil ist. Im Durchschnitt verdient ein Arbeitnehmer dieser Stichprobe 17,66 € (9,39)⁴⁴ brutto pro Stunde.

Bildungsdauer. Die Bildungsdauer ist vom DIW generiert, so dass pro allgemeinbildendem Abschluss stets die gleiche Anzahl an Bildungsjahren angenommen wird. Damit sollen Verzerrungen durch zum Beispiel Sitzenbleiber und Langzeitstudenten eliminiert werden. Einem Individuum, welches keinen Schulabschluss oder keine weiterführende Ausbildung hat, wird eine Gesamtzahl von 7 Bildungsjahren zugeordnet. Weiterhin werden bei vorhandenem Hauptschulabschluss 9, beim Realschulabschluss 10 und beim Abitur 13 Jahre zugeordnet. Je nach weiterer Qualifizierung ergeben sich dann Ergänzungen um 1,5 Jahre bei Abschluss einer einfachen, und um 2 Jahre bei Abschluss einer speziellen Lehrausbildung. Hiermit sind Ausbildungsberufe im Gesundheitsbereich gemeint. Für ein abgeschlossenes Fachhochschulstudium werden Ausbildungszeiten von 3 Jahren angesetzt, bei abgeschlossenem Hochschulstudium sind es 4 Jahre. Eine Weiterqualifizierung im akademischen Bereich findet keine eigene Berücksichtigung. Weiterhin wird bei Vorhandensein mehrerer Abschlüsse zur Berechnung der Bildungsdauer lediglich der höchste herangezogen. Die sich ergebenden möglichen Kombinationen von Bildungsdauern und deren zugehörige Abschlüsse sind in Tabelle 4.2 unter Einbezug der verschiedenen Stichprobenumfänge und der mittleren Bruttostundenlöhne näher aufgeschlüsselt. Im Durchschnitt hat ein Arbeitnehmer dieser Stichprobe eine Bildungsdauer von 12,8 Jahren (2,75).

Bildungsdauer versus Einkommen. Die theoretische Fundierung dieses Zusammenhangs ist in Abschnitt 2 über verschiedene Modelle dargestellt. Beim empirischen Vergleich der mittleren Bruttostundenlöhne pro Bildungsdauer wird der positive Zusammenhang beider Größen deutlich. Zur Veranschaulichung sind diese im Balkendiagramm in

⁴⁴Die in Klammern beschriebene Zahl stellt die Standardabweichung dar. Im Folgenden wird stets nachfolgend zum Mittelwert die Streuung in dieser Weise gezeigt.

4.2 Beschreibung der Variablen und deskriptive Analyse

Qualifikation	generierte Anzahl an Jahren	mittlerer Bruttolohn	Stichprobenumfang
Kein Schulabschluss	7	12,05 (4,61)	28
Ohne Hauptschulabschluss mit abgeschlossener Lehre	8,5	15,41 (5,79)	12
Ohne Hauptschulabschluss mit spezieller Ausbildung bzw. Hauptschulabschluss ohne Lehre	9	12,82 (4,72)	251
Realschulabschluss	10	13,33 (6,61)	114
Hauptschulabschluss mit Lehre	10,5	15,05 (5,65)	804
Hauptschulabschluss mit spezieller Ausbildung	11	16,64 (7,22)	219
Realschulabschluss mit Lehre	11,5	15,09 (7,23)	1031
Realschulabschluss mit spezieller Ausbildung	12	15,91 (7,58)	551
Abitur oder Realschule mit Fachhochschulreife	13	17,11 (8,57)	208
Fachhochschulreife und Lehre	13,5	18,33 (9,57)	82
Fachhochschulreife mit spezieller Ausbildung	14	19,81 (9,82)	75
Abitur und Lehre	14,5	17,47 (6,64)	142
Fachhochschulreife und Fachhochschulabschluss; Abitur und höher qualifizierende Ausbildungen	15	21,11 (12,26)	263
Abitur und Fachhochschulabschluss	16	21,52 (8,81)	208
Fachhochschulreife und Universitätsabschluss	17	28,54 (13,34)	19
Abitur und Universitätsabschluss	18	26,17 (12,34)	686

Tabelle 4.2: Anzahl der gesamt generierten Jahre mit zugehörigem Abschluss; Quelle: DIW, eigene Berechnungen.

4 Daten und Datenbeschreibung

Abbildung 4.5 dargestellt. Die Spannweite zwischen dem niedrigsten durchschnittlichen Einkommensniveau – bei keiner abgeschlossenen Ausbildung – und dem höchsten – bei einem abgeschlossenem Hochschulstudium – liegt bei 16,48 € pro Stunde. Dabei verdient ein Arbeitnehmer ohne Abschluss mit 12,05 € (4,61) am wenigsten. Der höchste Brutostundenlohn liegt bei 28,54 € (13,34) für Personen mit Fachhochschulreife und danach erworbenem universitärem Abschluss. Für die letztgenannte Personengruppe muss auf den geringen Stichprobenumfang hingewiesen werden. Die Standardabweichung der durchschnittlichen Einkommen schwankt⁴⁵ zwischen 4,61 € und 13,34 €, wobei die Schwankungsbreite bei Höherqualifizierten erwartungsgemäß deutlich größer ist.

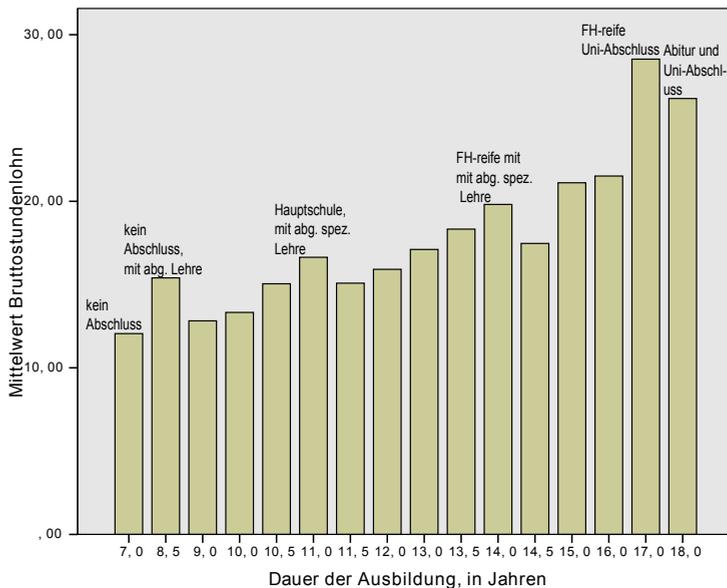


Abbildung 4.5: Dauer der Ausbildung versus mittlerer Bruttostundenlohn; Quelle: DIW, eigene Berechnungen.

Inwieweit sich die einzelnen Einkommensniveaus pro Bildungsjahr unterscheiden, wird nachfolgend behandelt. Hierfür wird ein zweiseitiger Zweistichprobentest⁴⁶ durchgeführt. Für diesen müssen zwei Annahmen erfüllt sein. Zum einen müssen die Teilgrundgesamt-

⁴⁵Dabei ist die geringste Standardabweichung bei jenen Personen ohne Abschluss, die höchste bei jenen, die im Mittel am meisten verdienen.

⁴⁶Die Vergleichsstichproben ergeben sich aus dem jeweiligen Vergleich einzelner Bildungslevel. Inwieweit das Einkommen verschieden ist, wird nun mit Hilfe eines Mittelwerttests zwischen den jeweiligen Stichproben untersucht. Dieser Test wird für alle Unterstichproben je nach Bildungsniveau angewandt.

heiten annähernd normalverteilt sein. Sachs und Hedderich (2006, 11. Auflage, S. 336), schlagen als grobe Voraussetzung vor, dass der Stichprobenumfang zumindest 60 Beobachtungen betragen sollte. Zum anderen gilt die Annahme der gleichen Varianzen pro Bildungsniveau. Die erste Annahme kann für fast alle Unterstichproben ab 9 Bildungsjahren bestätigt werden. Die Stichprobenumfänge bei 7, 8,5 und 17 Bildungsjahren sind dagegen zu gering. Die zweite Annahme kann nicht zwangsläufig angenommen werden, da davon auszugehen ist, dass mit steigendem Bildungslevel auch die Varianzen der mittleren Bruttostundenlöhne steigen. Um die Homogenität der Varianzen zu prüfen, wird der Levene-Test⁴⁷ angewendet. Die Hypothesen lauten:

$$H_0 : \sigma_i^2 = \sigma_j^2$$

$$H_1 : \sigma_i^2 \neq \sigma_j^2,$$

wobei σ^2 die Varianz der Teilgrundgesamtheit darstellt. Verglichen wird dabei die Varianz aus der i-ten und der j-ten Bildungsdauer, wobei $i \neq j$ gilt. Die Testergebnisse werden im Anhang A.2.2 dargestellt.

Kann die Nullhypothese bei einer Irrtumswahrscheinlichkeit von 5 % nicht abgelehnt werden, so liegt keine empirische Evidenz für eine deutliche Unterscheidung der Varianzen vor. Es wird daher in diesen Fällen angenommen, dass der t-Test angewendet werden kann. In Tabelle A.1 sind die P-Werte⁴⁸ für den Levene-Test dargestellt. Ist der angegebene P-Wert kleiner oder gleich 0,05, d. h. die Varianzen sind signifikant verschieden, so wird stattdessen der Welch-Test angewendet. Sowohl der t-Test als auch der Welch-Test vergleichen die sich ergebenden Mittelwerte. Letzterer korrigiert jedoch bei der Ermittlung der Teststatistik um die verschiedenen Varianzniveaus. In beiden Fällen lauten die Hypothesen für den zweiseitigen Mittelwertvergleich:

⁴⁷Näheres siehe Sachs und Hedderich (2006, 11. Auflage), S. 369.

⁴⁸Der P-Wert wird auch als Signifikanzwert bezeichnet und gibt an, mit welcher Wahrscheinlichkeit die Nullhypothese anzunehmen ist. Ist das Ziel, diese Hypothese zumindest mit einer Irrtumswahrscheinlichkeit von 5 % abzulehnen, so wird der kritische Wert bei 0,05 gesetzt. Alle niedrigeren P-Werte zeigen die Ablehnung der Hypothese bei gegebenem Signifikanzniveau. In den anderen Fällen wird die Nullhypothese nicht abgelehnt.

$$H_0 : \mu_i = \mu_j$$

$$H_1 : \mu_i \neq \mu_j,$$

wobei μ den Mittelwert der Teilgrundgesamtheit darstellt. Verglichen wird dabei das mittlere Einkommen aus der i -ten und der j -ten Bildungsdauer, wobei $i \neq j$ gilt.

In Tabelle 4.3 werden die jeweiligen P-Werte aus den Mittelwertvergleichen angegeben. Mit der gegebenen Irrtumswahrscheinlichkeit ist damit die Nullhypothese nicht signifikant ablehnbar, wenn der angegebene Wert $0,05$ übersteigt. In diesem Fall kann nicht bestätigt werden, dass die verglichenen mittleren Bruttostundenlöhne pro Bildungsdauer bei gegebener Sicherheit verschieden sind. Aus der Humankapitaltheorie von Becker (1994 (3. Auflage)) kann erwartet werden, dass die Nullhypothese mit zunehmender Unterschiedlichkeit der verglichenen Bildungsdauern auf dem 5 %-Niveau abzulehnen ist und demzufolge die mittleren Bruttostundenlöhne in ihrer Höhe nicht ähnlich sind. Aufgrund der geringen Stichprobenumfänge der Unterstichproben bei 7, 8,5 und 17 Bildungsjahren sind die Ergebnisse des Mittelwertvergleichs hierfür nur mit Vorsicht zu interpretieren.

Je nach Bildungsdauer unterscheiden sich die meisten mittleren Bruttostundenlöhne deutlich. Selbst bei geringen Bildungsunterschieden können schon signifikant verschiedene Bruttostundenlöhne ermittelt werden. Einige wenige Einschränkungen sind diesbezüglich jedoch vorzunehmen. So ist der Unterschied des Bruttostundenlohns zwischen 9 Bildungsjahren (mit Hauptschulabschluss ohne Ausbildung) und 10 Jahren (mit Realschulabschluss) nicht signifikant. Gleiches gilt im Vergleich von 10,5 Bildungsjahren (Hauptschulabschluss und abgeschlossene Lehre) und 11,5 Jahren (Realschulabschluss und abgeschlossene Lehre). Nach abgeschlossener Lehrausbildung erscheint es für den mittleren Bruttostundenlohn nicht relevant, ob ein Haupt- oder Realschulabschluss zugrundeliegt. Beim Vergleich der Löhne bei diesen Abschlüssen mit den sich ergebenden Löhnen nach Beendigung einer speziellen Ausbildung ergeben sich dagegen signifikante Unterschiede. Generell kann festgestellt werden, dass die speziellen Ausbildungen, also jene im Gesundheitsbereich, zu deutlich höheren Einkommen führen als „normale“ Lehrausbildungen, unabhängig vom vorherigen Schulabschluss. Der weitere Vergleich der mittleren Bruttostundenlöhne bei Real- oder Hauptschulabschluss und abgeschlossener spezieller Ausbildung führt wiederum zu keiner signifikanten Ablehnung der Gleichheitshypothese.

4.2 Beschreibung der Variablen und deskriptive Analyse

Bildungsdauer mitl. Brutto- std.- lohn	7	8,5	9	10	10,5	11	11,5	12	13	13,5	14	14,5	15	16	17	18
	28	12	251	114	804	219	1031	551	208	82	75	142	263	208	19	686
7	12,05	–														
8,5	15,41	0,06														
9	12,82	0,41	0,07													
10	13,33	0,27	0,38	–												
10,5	15,05	0,00	0,83	0,00	–											
11	16,64	0,00	0,56	0,00	0,00	–										
11,5	15,09	0,03	0,88	0,00	0,01	0,89	0,00	–								
12	15,91	0,00	0,82	0,00	0,02	0,23	0,03	–								
13	17,11	0,00	0,50	0,00	0,00	0,54	0,00	0,06	–							
13,5	18,33	0,00	0,31	0,00	0,00	0,10	0,00	0,01	0,29	–						
14	19,81	0,00	0,14	0,00	0,00	0,01	0,00	0,00	0,03	0,34	–					
14,5	17,47	0,00	0,30	0,00	0,00	0,27	0,00	0,03	0,67	0,43	0,07	–				
15	21,11	0,00	0,11	0,00	0,00	0,00	0,00	0,00	0,00	0,06	0,40	0,00	–			
16	21,52	0,00	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,17	0,00	0,69	–		
17	28,54	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00	0,01	0,04	–	
18	26,17	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,41	–

Tabelle 4.3: P-Werte aus den zweiseitigen Signifikanztests auf Gleichheit zwischen den mittleren Bruttostundenlöhnen pro Bildungsdauer; die kursiv geschriebenen Zahlen lassen die Ablehnung der Gleichheitshypothese nicht zu. Quelle: DIW, eigene Berechnungen.

Im Allgemeinen kann von signifikanten Einkommensunterschieden bei verschiedenen Bildungsdauern ausgegangen werden. Dabei gilt nicht nur, dass sich die höchste Bildungsdauer von der niedrigsten unterscheidet. Auch kleinere Bildungsunterschiede können meist als signifikant nachgewiesen werden. Demzufolge ist der generell ersichtliche Trend aus Abbildung 4.5 auch durch die zweiseitigen Zweistichprobentests nachzuweisen. Kleinere Schwankungen des mittleren Bruttostundenlohns entstehen durch eine unterschiedliche Bewertung beruflicher und allgemeiner Ausbildung.

Berufserfahrung. Mit dieser Variable wird die potentielle Dauer der Berufstätigkeit abgebildet. Dabei wird die Annahme getroffen, dass nach abgeschlossener allgemeiner Bildungsphase sofort die Jahre der Berufserfahrung anschließen. Der theoretisch fundierte Einfluss dieser Variablen zum Bruttostundenlohn wird aus dem in Abschnitt 2.2 beschriebenen erweiterten Modell nach J. Mincer ersichtlich. Die Erfahrungsvariable⁴⁹ berechnet sich aus

$$\text{Erfahrung} = \text{Alter} - \text{Bildungsdauer} - \text{Dauer der bisherigen Arbeitslosigkeit} - 6.$$

Es wird davon ausgegangen, dass mit 6 Jahren die allgemeinbildende Phase beginnt und das Individuum damit in die produktive Lebensphase einsteigt. Anders als bei anderen Untersuchungen⁵⁰ wird zusätzlich die Dauer der bisherigen Arbeitslosigkeit einbezogen, um der potentiellen Berufserfahrungsdauer⁵¹ besser gerecht zu werden.⁵² Als zusätzlicher Term wird auch das Quadrat der potentiellen Berufserfahrung integriert. Zum einen lässt sich dies aus dem theoretischen Modell von Mincer herleiten, welches in Abschnitt 2.2 dargestellt ist. Weiterhin konnte vielfach⁵³ empirisch ein quadratischer Zusammenhang nachgewiesen werden.⁵⁴ Inhaltlich bedeutet dieser Zusammenhang, dass das Grenzeinkommen mit zunehmender Zahl an Berufsjahren abnimmt. Dies bedeutet, dass das erste Jahr, in dem firmenspezifisches Humankapital akkumuliert wird, wesentlich produktivitätssteigernder wirkt als jedes weitere Jahr.

⁴⁹Hierbei ist zu berücksichtigen, dass in der generellen Bildungsphase kein firmenspezifisches Humankapital akkumuliert wird.

⁵⁰Siehe hierzu Mincer (1974), Lauer und Steiner (2000) und Ammermüller et al. (2009).

⁵¹Eventuelle Nichtarbeit in der produktiven Lebensphase (außer Arbeitslosigkeit), wie zum Beispiel durch Elternzeit, werden nicht berücksichtigt. Daher handelt es sich bei dieser Variablen um die potentielle Berufserfahrung.

⁵²Bei Kontrolluntersuchungen mit der von J. Mincer beschriebenen Berechnungsart finden sich nur marginal unterschiedliche Ergebnisse.

⁵³Siehe hierzu zum Beispiel Bennett et al. (1995), Rosen (1976) oder Hufnagel (2001).

⁵⁴Mincer (1974) diskutiert diesen Verlauf auch in seinem Buch, ab S. 64, ausführlich.

Geschlecht. In vielen Untersuchungen⁵⁵ wird der arbeitsmarktbezogene Einfluss des Geschlechts des Individuums betrachtet. In dieser Arbeit wird eine Kontrollvariable im Modell aufgenommen. Diese Dummy-Variable erhält den Wert 1, wenn das betrachtete Individuum weiblich ist. In dem gegebenen Datensatz sind dies 47 %. Der durchschnittliche Bruttostundenlohn der Männer liegt im Mittel bei 19,71 € (10,21) pro Stunde, Frauen verdienen hingegen⁵⁶ 15,37 € pro Stunde (7,77).

Familienstand. Die Einkommensvorteile von Verheirateten werden von Becker (1974) theoretisch analysiert. Auch empirisch wird durch die Berücksichtigung des Familienstandes ein positiver Effekt auf das Einkommen gemessen. Der Grund dafür kann sein „*dass Menschen, die in einer Ehe leben, umfangreichere soziale Netze knüpfen bzw. pflegen können als Alleinstehende oder Verwitwete.*“⁵⁷ Daher wird der Familienstand als eine Kontrollvariable in Form einer Dummy-Variablen in der Regression berücksichtigt. Die Ausprägung 1 bedeutet, dass eine Person verheiratet bzw. in einer eheähnlichen Gemeinschaft lebt. Alle anderen, also ledige, verwitwete, geschiedene oder Personen in Trennung werden mit 0 gelistet. 70 % der 4.693 in der Stichprobe berücksichtigten Personen sind verheiratet oder leben in einer eheähnlichen Gemeinschaft. Berechnet man für beide Gruppen getrennt den durchschnittlichen Bruttostundenlohn, so sprechen knapp 2 € pro Stunde Differenz⁵⁸ für die Nutzung von Spezialisierungsgewinnen oder umfangreicher sozialer Netze, die eine Ehe oder ein eheähnliches Zusammenleben ermöglichen. Es kann davon ausgegangen werden, dass Paare sich durch komparative oder absolute Vorteile privat sowie beruflich ergänzen und unterstützen. Im Mittel verdient eine verheiratete Person 18,22 € pro Stunde (9,82).

Unternehmensgröße. Je nach Art und Größe eines Unternehmens werden unterschiedliche Löhne gezahlt. So konnten Liu und Wu (2007) theoretisch sowie Brown und Medoff (1989) empirisch den Zusammenhang zwischen der Firmengröße und dem Einkommen fundieren. Dabei ist festzustellen, dass Mitarbeiter in großen Unternehmen tendenziell produktiver arbeiten. Dies ergibt sich zum Einen aufgrund der Ausnutzung von Skaleneffekten. Zum Anderen haben große Unternehmen eine größere Anziehungskraft, sodass sie

⁵⁵Siehe hierzu zum Beispiel Schnabel und Schnabel (2002).

⁵⁶Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.2 dargestellt.

⁵⁷Krenz (2008), S. 21.

⁵⁸Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.2 dargestellt.

im Mittel produktivere Arbeitnehmer für sich gewinnen können. Zu vermuten ist daher, dass Stundenlöhne in großen Unternehmen höher sind als in kleineren. Um dem gerecht zu werden, soll die Unternehmensgröße⁵⁹ in die Betrachtung einbezogen werden. Sie wird gemessen über die Anzahl der Mitarbeiter. In der Literatur gibt es keine einheitliche Abgrenzung bezüglich der Kategorisierung. Der Einfachheit halber wird die Unternehmensgröße als Dummy-Variable in der Regression berücksichtigt. Dabei ist ein Unternehmen groß, welches 200 und mehr Mitarbeiter beschäftigt. Individuen in großen Unternehmen erhalten den Wert 1. Ist ein Arbeitnehmer in einem Unternehmen mit weniger als 200 Mitarbeitern beschäftigt, so wird die Ausprägung der Dummy-Variable als 0 definiert. In der Stichprobe sind die Individuen gleichermaßen auf große und kleine Unternehmen verteilt. Erwartungsgemäß ist der durchschnittliche Stundenlohn mit 19,92 € (9,67) in großen Unternehmen signifikant⁶⁰ höher als jener in Unternehmen mit bis zu 200 Mitarbeitern. Sie verdienen 15,42 € pro Stunde (8,55).

Ländliches Gebiet. Die wirtschaftliche Struktur ist zwischen Stadt und Land heterogen. Alonso (1960) setzt sich mit den verschiedenen Strukturen zwischen urbanen und ländlichen Gebieten auseinander. Jochmann und Pohlmeier (2004) zeigen in ihrer Studie auch empirisch die Unterschiede auf. Daher wird eine weitere dichotome Variable in dieser Arbeit berücksichtigt. Dabei ist der Wohnort des Individuums die Entscheidungsbasis. Er ist ländlich, wenn er mehr als 25 km von einem Stadtzentrum entfernt ist. 61 % der Befragten wohnen im städtischen Gebiet. Der Mittelwertvergleich⁶¹ zeigt, dass Bewohner im ländlichen Raum weniger verdienen. Konkret liegt der Stundenlohn in der Stadt bei 18,16 € pro Stunde (9,45), in ländlichen Gegenden sind es mit 16,86 € pro Stunde (9,25) signifikant weniger.

Pendelfahrten. Pendelbewegungen nehmen heutzutage durch eine notwendig höhere Flexibilisierung weiter zu.⁶² Immer öfter wird die Wohnortentscheidung losgelöst vom Arbeitsort getroffen, beziehungsweise müssen mehrere Arbeitsstandorte in diese Überlegung einbezogen werden. Es kann von einer Selbstselektion ausgegangen werden. Es wird vermutet, dass diejenigen, die mehr Flexibilität zeigen, durch ein höheres Einkommen

⁵⁹Siehe hierzu auch Krenz (2008).

⁶⁰Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.2 dargestellt.

⁶¹Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.2 dargestellt.

⁶²Siehe hierzu Bogai et al. (2007).

entlohnt werden. Dies ist auch in der gegebenen Stichprobe zu beobachten. Es werden in der Analyse jene als Pendler angesehen, die mindestens wöchentlich pendeln müssen, sie erhalten den Wert 1. Allen Nichtpendlern wird die Ausprägung 0 zugeordnet. Zu dieser Gruppe gehören 56 % der Arbeitnehmer. Pendler verdienen im Mittel 18,38 € pro Stunde (9,79), die Nichtpendler dagegen verdienen mit 16,73 € pro Stunde (8,78) signifikant⁶³ weniger.

4.2.2 Variablen zur Schätzung der Bildungsdauer

Für eine bessere Schätzung der Bildungsrendite in Kapitel 5 ist es notwendig eine Zusatzregression vorzunehmen. Hierbei wird die Bildungsdauer bestimmt. Die möglichen Einflussfaktoren, die die Bildungsdauer, nicht aber das Einkommen bestimmen, werden in diesem Abschnitt dargestellt. Auch für dieses Regressionsmodell wird zunächst eine bivariate Analyse durchgeführt, die den Zusammenhang zwischen den möglichen Einflussfaktoren und der Bildungsdauer bestimmen.

Familiärer Hintergrund. Psacharopoulos (1977), Card (1995) und Dechênes (2007) zeigen den großen Einfluss der sozialen Herkunft auf. Eine besondere Bedeutung hat dabei der Einfluss des väterlichen Berufs auf die individuelle Bildungsdauer. Berücksichtigt wird die väterliche Bildung in der Zusatzregression. Dabei werden aus der vom DIW generierten kategorialen Schulvariable zwei Dummy-Variablen erstellt.

Die erste Variable steht für einen mittleren Abschluss, d. h. hat der Vater eines Individuums einen Realschulabschluss, das Abitur oder die Fachhochschulreife und eine abgeschlossene Berufsausbildung, so erhält die Variable mittlere Bildung vom Vater die Ausprägung 1, andernfalls 0. Die zweite Dummy-Variable nimmt den Wert 1 an, wenn der Vater ein abgeschlossenes Studium vorweisen kann. Dabei wird nicht zwischen einem Fachhochschul- und Universitätsabschluss unterschieden. Interpretiert werden die sich ergebenden Koeffizienten relativ zum niedrigsten Qualitätsniveau. Hierzu zählen all jene, die keinen Schulabschluss oder einen Hauptschulabschluss haben, ohne weitere abgeschlossene Berufsausbildung.

Durch die Verwendung von zwei Dummy-Variablen für die mittlere und hohe Bildung des Vaters werden die Koeffizienten relativ zum niedrigen Bildungslevel interpretiert. Auf-

⁶³Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.2 dargestellt.

grund dieser Dreigliedrigkeit der Kategorisierung der väterlichen Bildung ist es statistisch nicht möglich, Vater und Mutter gleichzeitig zu berücksichtigen. Die Interpretation der Koeffizienten würde miteinander kollidieren. Aufgrund des höheren Einflusses des Vaters⁶⁴ auf die Bildungsentscheidung des Kindes wird die kategoriale Bildungsvariable des Vaters⁶⁵ als Dummy mit Bezug zur untersten Bildungskategorie verwendet.

Der Vergleich der individuellen Ausbildungszeiten in Abhängigkeit vom Bildungsgrad des Vaters ergibt⁶⁶ für Personen, deren Väter ein geringeres Bildungsniveau haben, eine mittlere Bildungsdauer von 11,9 Bildungsjahren (1,96). Bei mittlerer väterlicher Bildung haben die Personen im Mittel 13,9 Bildungsjahre (3,18). In der letzten Gruppe sind es im Mittel sogar 17,0 Bildungsjahre (1,91).

Anzahl der Bücher. Anger et al. (2007) zeigen in einer Studie des Instituts der Deutschen Wirtschaft Köln (IW) empirisch den Zusammenhang zwischen der Bücheranzahl im Haushalt und der individuellen Bildungsdauer. Die Anzahl der Bücher wird vom DIW als nicht äquidistante, aber ordinalskalierte Variable zur Verfügung gestellt. Sie wird umkodiert in eine Dummy-Variable. Dabei wird die Grenze bei 200 Büchern gesetzt. All jene, die 200 und mehr Bücher besitzen, bekommen den Wert 1 zugeordnet. Alle anderen sind mit 0 gelistet. 69 % der Befragten haben weniger Bücher als der Grenzwert. Betrachtet man die mittleren Bildungsdauern separat, so ergibt sich für die Gruppe mit vielen Büchern eine mittlere Dauer von 14,4 Bildungsjahren (2,91). Die andere Gruppe hat mit im Mittel 12,1 Bildungsjahren (2,37) eine signifikant⁶⁷ geringere Qualifikation.

Anzahl der Geschwister. Es handelt sich bei der Anzahl der Geschwister um zwei metrische Variablen. Dabei werden die Anzahl der Brüder und Schwestern separat in die Regression einbezogen. Gemäß der Theorie von Willis (1973) besteht zwischen der Quantität und Qualität der Bildung der Kinder ein Substitutionsverhältnis, d. h. je mehr Kinder eine Familie hat, umso weniger Ressourcen können in die Qualität, also Ausbildung, eines jeden investiert werden. Der vermutete Zusammenhang zur Bildungsdauer nach Willis (1973) ist demnach – unabhängig vom Geschlecht der Geschwister – negativ.

⁶⁴Siehe hierzu Reilich (2006).

⁶⁵Berechnet man die Korrelationskoeffizienten der jeweiligen Schulbildung für die Dauer der Ausbildung, kann für beide Elternteile ein signifikanter Zusammenhang festgestellt werden. Der des Vaters ist jedoch mit 0,166*** deutlich größer, als jener der Mutter 0,041***.

⁶⁶Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.3 dargestellt.

⁶⁷Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.3 dargestellt.

4.2 Beschreibung der Variablen und deskriptive Analyse

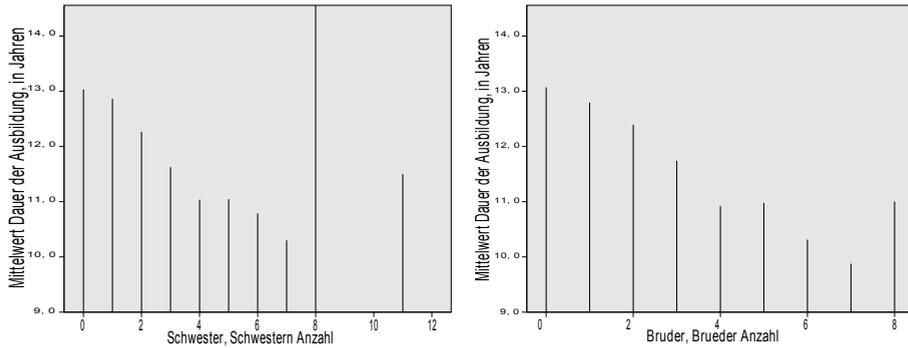


Abbildung 4.6: Mittlere Bildungsdauer pro Anzahl an Geschwistern; Quelle: DIW, eigene Berechnungen.

Auch empirisch wird der Zusammenhang durch Dechênes (2007) belegt. In Abbildung 4.6 ist diese Relation in Form von zwei Balkendiagrammen dargestellt. Dabei ist zu berücksichtigen, dass die Zahl der Beobachtungen mit steigender Zahl an Geschwistern exponentiell sinkt. So ist bei 8 und 11 Schwestern jeweils nur noch eine Beobachtung vorhanden. Mit zunehmender Zahl an Geschwistern sinkt wie erwartet das mittlere Bildungsniveau. Die Ausreißer bei 8 und 11 Schwestern sowie bei 8 Brüdern ist aufgrund der geringen Beobachtungszahl als unbedeutend zu werten.

Staatsangehörigkeit. Die Staatsangehörigkeit spielt durch verschiedene Kulturen, Werte und Normen eine bedeutende Rolle⁶⁸ für die Bildungsdauer. Die Dummy-Variablen Migration ist mit 0 gewertet, wenn das Individuum die deutsche Staatsbürgerschaft besitzt, andernfalls mit 1. In dem gegebenen Datensatz haben knapp 95 % der betrachteten Arbeitnehmer die deutsche Staatsbürgerschaft. Vergleicht man die Ausbildungslängen, so ergibt sich eine um 2 Jahre kürzere⁶⁹ mittlere Bildungsdauer für die Arbeitnehmer in Deutschland mit Migrationshintergrund. Ein Deutscher beschäftigt sich im Mittel knapp 13 Jahre (2,72) mit seiner Ausbildung.

⁶⁸Verschiedene Betrachtungen auf den Einfluss von Migration finden sich u. a. in Schwartz (1976) und Hunkler (2010).

⁶⁹Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.3 dargestellt.

Tätigkeit im erlernten Beruf. Die Dummy-Variable Tätigkeit im erlernten Beruf hat die Ausprägung 1, wenn die inhaltliche Tätigkeit sich seit der Ausbildung nicht verändert hat. 0 ergibt sich für all jene, die nicht mehr im Ausbildungsberuf tätig sind, bzw. gar keinen Beruf erlernt haben. 63 % der Befragten arbeiten im erlernten Beruf. Individuen, die ihre ausgeübte Tätigkeit zuvor erlernt haben, werden im Allgemeinen eine längere Ausbildung genossen haben. Das mag zum einen an verschiedenen Berufsklassen liegen – so eignen sich gering qualifizierte Berufe eher zum Quereinstieg – zum anderen hängt es auch von der Notwendigkeit der Substitution ab. Die Arbeitslosenquoten in höher qualifizierten Bereichen⁷⁰ belegen dies. Auch im verwendeten Datensatz wird der Zusammenhang deutlich. So haben jene, die die Tätigkeit nicht erlernt haben, im Mittel⁷¹ 11,8 Bildungsjahre (2,40), die andere Gruppe dagegen 13,3 Bildungsjahre (2,78).

Abschluss in der ehemaligen DDR. Um eine möglichst genaue Ermittlung der Bildungsdauer auch für die Personen zu erhalten, die im neuen Bundesgebiet die Schule besucht haben, ist es sinnvoll, für die verschiedenen politischen Gegebenheiten zu kontrollieren. Krueger und Pischke (1995) zeigen in ihrer Studie die Unterschiede in den Bildungsrenditen zwischen Ost- und Westdeutschland vor und nach der Wiedervereinigung. In dieser Arbeit ist eine Korrektur für die Lohngleichung nicht notwendig, da die Daten für das Jahr 2008 allgemein angeglichen wurden. Die Schulentcheidung jedoch ist zum Großteil schon zu Zeiten der innerdeutschen Teilung getroffen worden. Um dem approximativ gerecht zu werden, wird eine Dummy-Variable in die Regression einbezogen, die den Wert 1 annimmt, wenn die Person die Ausbildung in der ehemaligen DDR abgeschlossen hat. Die Ausprägung null steht dagegen für all jene, die einen westdeutschen Abschluss haben, bzw. für diejenigen, die zur Wendezeit noch in der Schule waren. Entscheidend ist dabei das Datum des Schulabschlusses. Von denen, die heutzutage im neuen Bundesgebiet wohnen, haben 88 % einen ostdeutschen Schulabschluss. Mit Blick auf die gewählte Altersspanne von 30–60 Jahren ist dies durchaus plausibel. In den alten Bundesländern leben 5 % mit einem Abschluss der ehemaligen DDR. Der Arbeitnehmer mit DDR-Abschluss hat im Mittel mehr Zeit mit seiner Ausbildung verbracht. So liegt die durchschnittliche Ausbildungsdauer⁷² bei 13,1 Jahren (2,53), die Arbeitnehmer mit Westabschluss besuchten qualifizierende Einrichtungen im Schnitt 12,7 Jahre (2,81) lang. Dies erscheint auf den

⁷⁰Siehe dazu empirisch Steiner und Schmitz (2010) und theoretisch Micklewright et al. (1990).

⁷¹Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.3 dargestellt.

⁷²Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.3 dargestellt.

ersten Blick kontraintuitiv. Begründung dafür könnte der Umbruch durch die Wendezeit sein. Es ist zu vermuten, dass sich diejenigen mit höherem Abschluss eher in die westdeutsche Gesellschaft integrieren konnten. Daher muss davon ausgegangen werden, dass die Hochqualifizierten überproportional vertreten sind.

4.2.3 Separierende Variablen des Datensatzes

In diesem Abschnitt werden Häufigkeitsverteilungen sowie Mittelwerte für ausgewählte Merkmale verglichen. Dabei wird im Besonderen auf Altersgruppen, den öffentlichen Dienst im Vergleich zur Privatwirtschaft sowie Voll- und Teilzeitbeschäftigte eingegangen.

Kohorten. Die Entscheidung für oder gegen einen Ausbildungsweg ist immer auch im Kontext der aktuellen politischen und gesellschaftlichen Gegebenheit zu verstehen. Um dem Rechnung zu tragen, werden vielfach bei Untersuchungen zur Bildungsrendite verschiedene Kohorteneffekte berücksichtigt. So verwenden Schnabel und Schnabel (2002) in der Lohnregression eine Kohorten-Kontrollvariable. Steiner und Lauer (2000) dagegen separieren den Datensatz nach bestimmten Altersgruppen und schätzen verschiedene Renditen pro Kohorte.

In dieser Arbeit soll in Anlehnung an letztgenannte Autoren eine Aufteilung des Datensatzes erfolgen. Hierfür wird dieser gedrittelt, so dass die erste Gruppe diejenigen beinhaltet, die vor 1958 geboren sind. Die Personen der zweiten Gruppe sind zwischen 1958 und 1967 geboren und die der dritten Gruppen zwischen 1968 und 1978. Alle späteren bzw. früheren Generationen werden nicht berücksichtigt, da die Stichprobe auf die 30–60-Jährigen reduziert wurde.

Altersspanne	n	relative Verteilung zum gesamten Datensatz	mittlere Bildungsdauer in Jahren	mittlerer Bruttostundenlohn in €
30–40	1410	30 %	12,9 (2,75)	15,97 (7,34)
41–50	1871	40 %	12,6 (2,64)	17,88 (9,33)
51–60	1412	30 %	12,8 (2,89)	19,05 (10,93)

Tabelle 4.4: Deskriptive Statistiken pro Kohorte; Quelle: DIW, eigene Berechnungen.

Tabelle 4.4 zeigt die relative Häufigkeitsverteilung dieser drei Gruppen. Die Anteile sind für die jüngste und älteste Kohorte gleich. Die mittlere Altersklasse ist mit 40 % mit deutlich höherem Anteil vertreten als die anderen beiden. In dieser Tabelle werden zusätzlich die sich pro Kohorte ergebenden Mittelwerte für den Bruttostundenlohn und die mittlere Bildungsdauer angegeben. Es wird ersichtlich, dass die Bildungsdauern⁷³ für die Kohorten ungefähr konstant sind. Ein leichter Rückgang der Bildungsjahre ist im mittleren Alterssegment zu finden. Anders hingegen der Bruttostundenlohn. Die intuitiv vermutete altersgemäße Steigerung des Einkommens wird durch die mittleren Bruttostundenlöhne bestätigt. Um den reinen Einfluss der Bildung pro Alterssegment zu ermitteln, ist eine Regressionsanalyse erforderlich, wobei für die Dauer der Berufserfahrung kontrolliert⁷⁴ werden muss.

Öffentlicher Dienst. Steiner und Lauer (2000) konnten eine höhere Rendite im öffentlichen Sektor nachweisen. Die Frage, ob der jeweilige Arbeitnehmer dieses Datensatzes im öffentlichen Dienst beschäftigt ist oder nicht, wird nicht auf jährlicher Basis erhoben. Lediglich diejenigen, die ihre Arbeitsstelle gewechselt haben, wurden im Jahr 2008 um diese Angabe gebeten. Für alle anderen werden die Angaben aus dem Jahr 2007⁷⁵ übernommen. 30 % der Arbeitnehmer waren im öffentlichen Dienst tätig.

Ein Arbeitnehmer des öffentlichen Dienstes hat im Mittel⁷⁶ eine durchschnittliche Bildungsdauer von 13,6 Jahren (2,94). Diese Anzahl an Bildungsjahren liegt 1,3 Jahre über dem Durchschnitt der Nichtselbstständigen, die im privaten Sektor tätig sind. Auch der Stundenlohn im öffentlichen Dienst liegt über dem der Privatwirtschaft. So verdient ein im öffentlichen Dienst Tätiger 18,73 € pro Stunde (8,18), wohingegen die Privatwirtschaft im Durchschnitt lediglich 17,22 € (9,87) zahlt. Intuitiv wird vermutet, dass die Bruttolöhne in der Privatwirtschaft über denen des öffentlichen Dienstes liegen um u. a. dem höheren Risiko der Arbeitslosigkeit gerecht zu werden. Betrachtet man hingegen die Standardabweichungen beider Bruttostundenlöhne, so kann festgestellt werden, dass die mittlere quadratische Abweichung in der Privatwirtschaft deutlich größer ist. Auch aus der gegebenen durchschnittlichen Bildungsdauer kann geschlussfolgert werden, dass sich die Bildungsstruktur zwischen dem öffentlichen Dienst und der Privatwirtschaft unter-

⁷³Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.4 dargestellt.

⁷⁴Näheres siehe Abschnitt 5.

⁷⁵Im Jahr 2007 hatten alle Teilnehmer am SOEP diese Frage zu beantworten.

⁷⁶Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.4 dargestellt.

scheidet. Um eine bessere Vergleichbarkeit herzustellen, sind in Tabelle 4.5 die mittleren Bruttostundenlöhne für drei ausgewählte Bildungsdauern dargestellt.

Unterschiede in der Entlohnung bei Personen mit Hauptschulabschluss oder Realschulabschluss und jeweils abgeschlossener Lehrausbildung sind zwischen dem öffentlichen und privaten Sektor nicht signifikant⁷⁷. Im Gegensatz dazu unterscheidet sich der mittlere Bruttostundenlohn für Personen mit einem Universitätsabschluss zwischen öffentlichem und privaten Sektor besonders deutlich.

ausgewählter Bildungsabschluss	gen. Bildungsdauer	Bruttostundenlohn in €			
		öffentlicher Dienst	Anteil	Privatwirtschaft	Anteil
Hauptschulabschluss mit Lehre	10,5	14,82 (4,81)	0,103	15,14 (5,80)	0,200
Realschulabschluss mit Lehre	11,5	15,94 (4,78)	0,181	14,82 (7,88)	0,238
Hochschulabschluss	18	25,23 (9,77)	0,245	27,16 (14,47)	0,107

Tabelle 4.5: Vergleich der Bruttostundenlöhne im privaten und öffentlichen Sektor für ausgewählte Bildungsdauern; Quelle: DIW, eigene Berechnungen.

Ein Mitarbeiter des öffentlichen Dienstes verdient somit im Mittel 25,23 € pro Stunde (9,77). Der Arbeitnehmer mit diesem Abschluss, tätig in der Privatwirtschaft, verdient dagegen mit im Mittel 2 € signifikant mehr. Die vermutet unterschiedliche Bildungsstruktur zwischen öffentlichem Dienst und Privatwirtschaft ist durch die relativen Häufigkeiten pro Sektor zu bestätigen. So sind die meisten im öffentlichen Dienst Beschäftigten hochqualifiziert. Im Gegensatz dazu ist die modale Klasse der Privatwirtschaft die mit mittlerer Qualifikation, dicht gefolgt von jenen mit Hauptschulabschluss und Lehrausbildung. Diese drei Bildungsdauern stellen lediglich eine kleine Auswahl dar und verdeutlichen die anfängliche Vermutung unterschiedlicher Bildungsstrukturen in beiden Bereichen.

Voll- und Teilzeitbeschäftigung. Lauer und Steiner (2000) haben in ihrer Untersuchung mit den SOEP-Daten aus den Jahren 1984 bis 1997 ermitteln können, dass die

⁷⁷Die Testergebnisse aus dem Levene- und Mittelwertvergleichstests für die ausgewählten Bildungsdauern sind im Anhang A.2.2 in Tabelle A.5.

Bildungsrendite zwischen voll- und teilzeitbeschäftigten Arbeitnehmern stark schwankt. Gemäß der Untersuchung erhielten Teilzeitbeschäftigte eine wesentlich höhere Bildungsrendite als jene, die in Vollzeit arbeiten.

Für die hier anstehende Untersuchung gelten Individuen als teilzeitbeschäftigt, wenn sie nicht mehr als 30 Stunden pro Woche vertraglich vereinbart haben. In der Stichprobe sind knapp 24 % teilzeitbeschäftigt. Die mittlere Bildungsdauer liegt bei beiden Teilgruppen bei ca. 12,7 Jahren (2,7). Der durchschnittliche Stundenlohn der Teilzeitbeschäftigten liegt dagegen mit 15,65 € (9,09) signifikant⁷⁸ niedriger, als derer, die mehr als 30 Stunden arbeiten. Sie verdienen durchschnittlich 18,28 € pro Stunde (9,40). Damit kann trotz gleichem Bildungsniveau ein recht deutlicher Unterschied von ca. 2,50 € im mittleren Bruttostundenlohn zugunsten der Vollzeitbeschäftigten ermittelt werden.

4.2.4 Zusammenfassung

Dieser Abschnitt diskutiert die nachfolgend verwendeten Variablen und motiviert ihre Verwendung in dem gewählten Zusammenhang. Dabei werden zunächst die beiden für diese Arbeit bedeutenden Variablen ausführlich diskutiert und miteinander in Relation gesetzt. Aus dem reinen Vergleich von Mittelwerten wird deutlich: Bildung beeinflusst das Einkommen. Weitere Variablen, die zur Erklärung des Bruttostundenlohns bzw. der Bildungsdauer verwendet werden können, sind diskutiert und durch weitere Literatur fundiert. Ebenfalls wird die separierte Analyse nach drei ausgewählten Merkmalen begründet.

4.3 Korrelationsbeziehungen

In diesem Abschnitt werden bivariate Korrelationsbeziehungen diskutiert. Wie im Abschnitt 4.2 sind dabei die erklärenden Variablen für den Bruttostundenlohn und die Bildungsdauer berücksichtigt.

Generell stellen Korrelationskoeffizienten den bivariaten Zusammenhang zwischen zwei Merkmalen dar. Das Intervall dieser dimensionslosen Kennzahl liegt zwischen -1 bei einem stark negativen und $+1$ bei einem stark positiven Zusammenhang. Ist der Wert bei 0 , so ist von keinem Zusammenhang zwischen beiden Merkmalen auszugehen. Zur Berech-

⁷⁸Die Testergebnisse aus Levene- und zweiseitigem Zweistichproben Mittelwerttest sind im Anhang A.2.2 in Tabelle A.4 dargestellt.

nung der Korrelationskoeffizienten wird für metrisch skalierte oder dichotome Merkmale der Koeffizient von Bravais und Pearson verwendet. Dieser misst den linearen statistischen Zusammenhang zwischen zwei Merkmalen.⁷⁹ Voraussetzung ist eine lineare Beziehung zwischen beiden Merkmalen, bei der zudem beide Merkmale annähernd normalverteilt sind. Die Signifikanzprüfung erfolgt über einen t-Test.

Der Korrelationskoeffizient von Bravais und Pearson liefert keine Aussage über die Richtung des Zusammenhangs. Weiterhin kann „*nicht zwingend von einer kausalen Beziehung zwischen den Merkmalen*“⁸⁰ ausgegangen werden, sondern lediglich von einer statistischen Tendenz. Es ist nicht auszuschließen, dass ein möglicher Zusammenhang durch eine Scheinkorrelation gegeben ist. Dabei suggeriert der Korrelationskoeffizient einen positiven oder negativen Zusammenhang, tatsächlich jedoch ist eine dritte, unberücksichtigte Größe mit beiden Merkmalen korreliert. Die Verwendung dieses Maßes ist daher nur als ein Indiz für einen möglichen Zusammenhang zu verstehen. Er dient dazu, einen weiteren Überblick über den Datensatz zu erhalten und Beziehungen zum Bruttostundenlohn bzw. zur Bildungsdauer herzustellen. Diese werden in der Regressionsanalyse in Abschnitt 5.1 vertieft.

4.3.1 Korrelationsbeziehungen zum Bruttostundenlohn

Die Korrelationsbeziehungen zwischen dem Bruttostundenlohn und den bereits in Abschnitt 4.2.1 beschriebenen Variablen werden in Tabelle 4.6 dargestellt.

Variable	Korrelationskoeffizient zum logarithmierten Bruttostundenlohn
Bildungsdauer	0,421***
Erfahrung	0,037***
Quadrierte Erfahrung	0,026*
Geschlecht	-0,251***
Familienstand	0,086***
Unternehmensgröße	0,287***
ländliches Gebiet	-0,082***
Pendler	0,097***

Tabelle 4.6: Korrelationen zum logarithmierten Stundenlohn; Quelle: DIW, eigene Berechnungen. Signifikanzniveau: *** 1 %, ** 5 % und * 10 %.

⁷⁹Die zugrundeliegende Berechnungsformel lautet: $r_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$, wobei r das Zusammenhangsmaß zwischen den Merkmalen X und Y beschreibt.

⁸⁰Schira (2005), S. 94.

Verwendet wird an dieser Stelle der logarithmierte⁸¹ Bruttostundenlohn, da hierbei eine bessere Vergleichbarkeit mit der Regression in Abschnitt 5 erreicht wird. Der aus Abschnitt 4.2 vermutete positive Zusammenhang zwischen der Bildungsdauer und dem Bruttostundenlohn kann auch auf Basis der Korrelationskoeffizienten gezeigt werden. Die Berufserfahrung wirkt lediglich in einfacher Form signifikant positiv auf das Einkommen. Das mag daran liegen, dass für die Berechnung des Korrelationskoeffizienten nach Bravais und Pearson zwingend ein linearer bivariater Zusammenhang angenommen wird, dieser jedoch hier nicht⁸² zugrunde liegt. Die Berücksichtigung dessen erfolgt innerhalb der Regressionsanalyse in Abschnitt 5. Der negative Zusammenhang zwischen dem weiblichen Geschlecht und dem logarithmierten Bruttostundenlohn ist auch mit diesem Korrelationsmaß signifikant nachweisbar. Gleiches gilt für den positiven Einfluss des Familienstandes. Der Zusammenhang zwischen dem Einkommen und der Unternehmensgröße ist im Vergleich zum Familienstand deutlich stärker. In einem großen Unternehmen wird tendenziell ein höheres Einkommen erzielt. Ist der Wohnort eines Arbeitnehmers im ländlichen Gebiet, so hat er eher ein geringeres Einkommen als jener der im städtischen Raum lebt. Zeigt der Arbeitnehmer die Bereitschaft zumindest einmal pro Woche zu pendeln, so hat er tendenziell ein höheres Einkommen.

4.3.2 Korrelationsbeziehungen zur Bildungsdauer

Der Zusammenhang zwischen der Bildungsdauer und den in Abschnitt 4.2.2 dargestellten Variablen wird im nächsten Schritt über den Korrelationskoeffizienten von Bravais und Pearson gemessen. Der Zusammenhang zwischen der mittleren bzw. hohen väterlichen Bildung und der individuellen Bildungsdauer ist erwartungsgemäß durch einen hohen und signifikanten Koeffizienten zu bestätigen. Auch die Anzahl der Bücher im Haushalt, ein Schulabschluss aus der ehemaligen DDR⁸³ und die aktuelle Tätigkeit im erlernten Beruf zeigen einen positiven Zusammenhang zur Bildungsdauer. Gemäß der bereits in Abschnitt 4.2.2 erwähnten Theorie von Willis (1973) kann auch über die Korrelationskoeffizienten ein negativer Zusammenhang zwischen der Zahl der Geschwister und der individuellen

⁸¹Der Vergleich der Ergebnisse mit dem einfachen Bruttostundenlohn ergibt marginale Abweichungen, die die Tendenz der Aussage nicht verändern.

⁸²Siehe hierzu u. a. Abschnitt 4.2.1.

⁸³Dieses zunächst den Erwartungen widersprechende Ergebnis kann mit Blick auf die Erwerbsstrukturen relativiert werden. Nach der Wende haben sich zumeist diejenigen mit einer höheren Qualifikation auf dem Arbeitsmarkt behaupten können.

Bildungsdauer gezeigt werden. Erwerbstätige ohne deutschen Pass haben tendenziell eine geringere Bildungsdauer.

Variable	Korrelationskoeffizient zu Bildungsdauer
mittlere väterliche Bildung	0,267***
hohe väterliche Bildung	0,392***
viele Bücher	0,377***
Zahl der Schwestern	-0,154***
Zahl der Brüder	-0,165***
Schulabschluss aus der ehem. DDR	0,068***
Tätigkeit im erlernten Beruf	0,279***
Staatsangehörigkeit nicht deutsch	-0,164***

Tabelle 4.7: Korrelationen zur Bildungsdauer; Quelle: DIW, eigene Berechnungen. Signifikanzniveau: *** 1 %, ** 5 % und * 10 %.

4.3.3 Zusammenfassung

Die bereits motivierten Variablen aus dem vorangegangenen Abschnitt werden nicht nur über Mittelwertvergleiche gegeneinander getestet. Bevor eine multivariate Schätzung Zusammenhänge aufdeckt, wird zunächst der bivariate Zusammenhang für diese Stichprobe untersucht. Für alle identifizierten Variablen bestätigt sich der vermutete Zusammenhang.

5 Schätzmethoden und Ermittlung der nationalen Bildungsrendite

In diesem Abschnitt werden verschiedene Methoden und Modelle zur Schätzung der Bildungsrendite untersucht. Dafür werden die unterschiedlichen Methoden zunächst beschrieben und sinnvolle Verfahren werden identifiziert. Ein Vergleich verschiedener Regressionsmodelle hat im nächsten Schritt das Ziel, jenes mit dem größten Erklärungsgehalt zu ermitteln. Als Ergebnis soll eine unverzerrte nationale Rendite geschätzt werden. Dann erfolgt eine Separierung für verschiedene Bevölkerungsgruppen mit dem Vergleich der ermittelten Renditen. Im Weiteren wird die Möglichkeit der Erweiterung des Datensatzes um die Gruppe der Erwerbslosen diskutiert. Zunächst wird dafür der Datensatz um die ALG-I-Empfänger erweitert und Renditen werden auf Basis der erweiterten sowie zweier separierter Stichproben geschätzt. Im nächsten Schritt wird das Heckman-Verfahren diskutiert. Mit diesem ist es möglich alle Erwerbspersonen in die Analyse einzubeziehen. Hierfür wird der Datensatz um die noch fehlende Gruppe erweitert und auf Repräsentativität geprüft. Die Schätzung nationaler Bildungsrenditen mit diesem Verfahren folgt.

5.1 Schätzverfahren – Methode der kleinsten Quadrate (OLS)

Ziel der Regressionsanalyse⁸⁴ ist es, erwartungstreue, konsistente und effiziente Schätzer⁸⁵ für die Messung des Zusammenhangs zwischen den erklärenden und der zu erklärenden Variable zu ermitteln. Unter Erwartungstreue wird verstanden, dass der tatsächliche Wert approximativ dem geschätzten Wert des Schätzers entspricht und dieser damit unverzerrt

⁸⁴Dieser Abschnitt basiert auf Backhaus et al. (2006 (11. Auflage)), Eckey et al. (2004), Heij et al. (2004) und Wooldridge (2002).

⁸⁵Dieser Best Linear Unbiased Estimator (BLUE) errechnet sich aus $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$, wobei X der Vektor des Regressors und y die abhängige Variable beschreibt.

ist. Konsistent ist ein Koeffizient immer dann, wenn bei Vergrößerung des Stichprobenumfangs der Schätzer dem wahren Wert weiter entgegen strebt. Ist ein Schätzer effizient, so bedeutet dies, dass er „unter allen linearen und unverzerrten Schätzern eine kleinstmögliche Varianz“⁸⁶ aufweist. Um dies für die OLS-Regression zu gewährleisten, müssen einige Voraussetzungen erfüllt sein. Sie werden im Folgenden dargestellt werden.⁸⁷

Voraussetzungen zur Anwendung des OLS-Methoden.

V1 *Das Regressionsmodell ist richtig aufgestellt, d. h. es enthält alle zur Erklärung nötigen Variablen. Weiterhin muss ein linearer Zusammenhang zwischen Schätzern und Regressand bestehen. Zudem muss die Schätzbarkeit des Modells gewährleistet sein. Für ein solches Modell muss grundsätzlich gelten, dass „die Zahl der zu schätzenden Parameter $(J+1)$ [...] kleiner [ist,] als die Zahl der vorliegenden Beobachtungen“⁸⁸.*

Ein perfektes Regressionsmodell enthält alle relevanten Faktoren. In der Empirie kann man sich diesem häufig nur annähern, da zumeist nicht alle Einflüsse identifizierbar und quantifizierbar sind. Ist das Modell nicht perfekt spezifiziert, kann es zu Verzerrungen der Koeffizienten kommen. Sind im Modell zu viele Variablen berücksichtigt (Overfitting), so führt dies dazu, dass eigentlich relevante Faktoren als irrelevant gekennzeichnet werden. Im gegenteiligen Fall, beim Underfitting fehlen tatsächlich relevante Variablen. Dabei muss von einer Korrelation zwischen den Regressoren und dem Störterm⁸⁹ ausgegangen werden.

Zur Identifikation des besten Modells wird ein Vergleich verschiedener Spezifikationen in Abschnitt 5.1.1 vorgenommen. Dabei wird der Erklärungsgehalt des jeweiligen Modells näher untersucht sowie die Stabilität der einzelnen Koeffizienten aufgezeigt. Verändern sich die Koeffizienten stark bei Hinzunahme von Regressoren, deutet dies auf eine falsche Spezifikation hin. Im Fokus steht die Schätzung der Bildungsrendite. Demzufolge wird bei dem Modellvergleich besonders auf die Stabilität des Schätzers für die Bildungsrendite Wert gelegt. Problematisch ist dabei die mangelnde Quantifizierbarkeit der individuell verschiedenen Motivationen und Fähigkeiten. Dieses Problem wird in Abschnitt 5.2 näher betrachtet.

⁸⁶Backhaus et al. (2006 (11. Auflage)), S. 79.

⁸⁷Die Annahmen sowie die zugehörigen Testverfahren sind zusammengefasst im Anhang A.3.1.

⁸⁸Backhaus et al. (2006 (11. Auflage)), S. 79.

⁸⁹Siehe hierzu auch V4.

5.1 Schätzverfahren – Methode der kleinsten Quadrate (OLS)

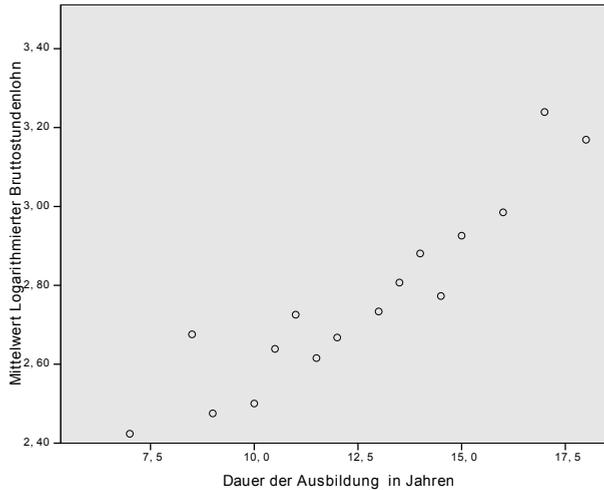


Abbildung 5.1: Vergleich der Dauer der Ausbildung mit dem mittleren logarithmierten Brutostundenlohn; Quelle: DIW, eigene Berechnungen.

Die Annahme der Linearität bezieht sich auf den Zusammenhang⁹⁰ zwischen Regressand und Koeffizient. Dies wird in den Modellspezifikationen berücksichtigt. In Abschnitt 2 ist theoretisch ein linearer Zusammenhang zwischen dem logarithmierten Einkommen und der Bildungsdauer hergeleitet worden. Abbildung 5.1 ermöglicht die empirische Bestätigung dieses linearen Zusammenhangs. Die generelle Schätzbarkeit ist wegen der großen Beobachtungszahl nicht kritisch.

V2 Die Störgröße ε_i muss eine stochastische und damit nichtsystematische, normalverteilte Komponente sein. Dann gilt, dass $E(\varepsilon_i) = 0$ ist.

Ist die Bedingung nicht erfüllt, so ist von einem systematischen Messfehler auszugehen.⁹¹ Dies kann zur Verzerrung der Konstanten führen. Häufig ist diese lediglich ein zusätzlicher Term, der nicht von eigentlichem Interesse ist. Demzufolge hat eine verzerrte Schätzung keinen negativen Einfluss auf die eigentlich interessierenden Koeffizienten. Die theoretische Herleitung in Abschnitt 2.2 zeigt die vielfältigen Einflussgrößen auf den absoluten Term in der Lohngleichung. Eine Interpretation wäre auf Basis dieser Erkenntnis ohnehin schwierig. Der Fokus dieser Arbeit liegt auf dem Zusammenhang zwischen Bildung und Einkommen. Demnach spielt der

⁹⁰Eine mögliche Spezifikation in der Form: $Y_i = \alpha \cdot \prod_{j=0}^J X_j^{\beta_j} \cdot \varepsilon_i$ ist nicht schätzbar.

⁹¹Siehe hierzu Backhaus et al. (2006 (11. Auflage)).

absolute Term keine entscheidende Rolle und eine Verzerrung dieses Schätzers hätte keinen Einfluss auf die Aussagekraft für die zu schätzende Bildungsrendite.

Die Annahme der normalverteilten Störgröße ist eine Zusatzannahme. Ist sie nicht erfüllt, so können Schätzer trotzdem den oben beschriebenen⁹² Kriterien entsprechen. Problematisch hingegen wird die Durchführung statistischer Tests. Bei diesen wird unterstellt, dass die geschätzten Koeffizienten einer Normalverteilung folgen. Durch die Linearkombination gilt, dass, „wenn die Störgrößen normalverteilt sind, dann sind auch die Y -Werte, die die Störgröße als additiven Term enthalten, normalverteilt“⁹³ und damit gilt diese Annahme ebenso für die geschätzten Koeffizienten. Der Jarque-Bera-Test prüft die Nullhypothese der Normalverteilung.⁹⁴ Muss die Nullhypothese für die Störgröße verworfen werden, so können die Schätzer für die Regressionskoeffizienten trotzdem der geforderten Verteilung folgen. Asymptotisch gilt für eine große Zahl⁹⁵ an Beobachtungen, dass „die Signifikanztests unabhängig von der Verteilung der Störgröße gültig“⁹⁶ sind. Die Verteilung der Störgröße muss für jede einzelne Modellspezifikation geprüft werden. Die Ergebnisse des Jarque-Bera-Tests finden sich in den Ergebnistabellen jeweilig pro Modellspezifikation.

V3 *Es liegt Homoskedastizität vor, d. h. es wird von einer für alle Elemente konstanten Varianz ausgegangen, formal bedeutet dies: $\text{Var}(\varepsilon_i) = \sigma^2$.*

Bei Vorhandensein von Heteroskedastizität sind die aus der Regression geschätzten Koeffizienten weiterhin erwartungstreu und konsistent. Jedoch ist die Effizienz nicht gegeben. Mit dem Breusch-Pagan-Test⁹⁷ wird die Nullhypothese der Homoskedastizität geprüft. Muss sie verworfen werden, so sind die Signifikanzniveaus nicht belastbar und damit besitzen die statistischen Tests keine Gültigkeit. Bei Ablehnung

⁹²Siehe die Kriterien für BLUE zu Beginn dieses Abschnitts.

⁹³Backhaus et al. (2006 (11. Auflage)), S. 92.

⁹⁴Zur Berechnung der JB-Wertes wird folgende Formel angewendet: $JB = n[\frac{1}{6}S^2 + \frac{1}{24}(K - 3)^2]$, wobei S für die Schiefe und K für die Kurtosis der Verteilung steht.

⁹⁵Backhaus et al. (2006 (11. Auflage)), S. 94 schlägt hierbei einen Mindeststichprobenumfang von 40 Beobachtungen vor.

⁹⁶Backhaus et al. (2006 (11. Auflage)), S. 93.

⁹⁷Breusch-Pagan-Lagrange-Multiplier-Test, siehe hierzu Heij et al. (2004). Die sich aus der Regression ergebenden Residuen werden in quadrierter Form als abhängige Variable auf die unabhängigen Variablen des ursprünglichen Modells geschätzt. Bei großen Stichprobenumfängen kann näherungsweise der Lagrange-Multiplikator mit $LM_{BP} = nR^2$ berechnet werden, wobei das R^2 aus der Zusatzregression stammt.

der Homoskedastizität schaffen die Standardfehler von Huber, White und Eicker⁹⁸ Abhilfe. Die sich ergebenden White-Standardfehler sind robust gegen Heteroskedastizität. Der P-Wert des Breusch-Pagan-Tests wird pro geschätztem Modell angegeben. Ist die Nullhypothese abzulehnen, sind die angegebenen Standardfehler die nach White bereinigten.

- V4** *Die Regressoren sind exogen, d. h. es besteht keine Korrelation zwischen dem Störterm und den unabhängigen Variablen. Formal muss damit gelten: $E(x_{ij}, \varepsilon_i) = 0$.*

Endogenität „describe[s] any situation where an explanatory variable is correlated with the disturbance“⁹⁹. Ist der Erwartungswert größer als null, so wird der Koeffizient β_j überschätzt und umgekehrt. Dieses Problem kann durch fehlende Variablen oder Messfehler entstehen.¹⁰⁰ Diese Voraussetzung steht in starkem Zusammenhang zur in V1 geforderten richtigen Modellspezifikation. Der Test von Durbin, Wu und Hausmann¹⁰¹ prüft die Nullhypothese der Exogenität. Ist die Nullhypothese abzulehnen, so zeigt dies, dass die Schätzer möglicherweise inkonsistent, ineffizient und nicht erwartungstreu sind. Die Darstellung der Testergebnisse erfolgt bei der jeweiligen Modellspezifikation in der Ergebnisdarstellung. Als Vergleich wird eine Instrumentvariablenschätzung benötigt. In Abschnitt 5.2.1 werden verschiedene Modellspezifikationen zur richtigen Instrumentierung betrachtet. Das beste Modell¹⁰² wird identifiziert und gilt als Referenzmodell für diesen Test.

- V5** *Die Störterme sind nicht korreliert. Es gilt: $Cov(\varepsilon_i, \varepsilon_{i+r}) = 0$, wenn $r \neq 0$. Es besteht damit keine Autokorrelation.*

Häufig besteht das Problem der Autokorrelation bei Zeitreihen.¹⁰³ So ist beispiels-

⁹⁸Hierfür wird die Varianz geschätzt über $V\hat{A}R(\beta_j) = \frac{\sum \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$, wobei \hat{u}_i für die Residuen der ursprünglichen OLS-Regression, \hat{r}_{ij} für die Residuen aus der Zusatzregression (näheres siehe Anhang A.3.1) steht. SSR_j ist die Summe der Residuen dieser Zusatzregression. Die Wurzel aus der Varianz ergibt die White-Standardfehler. Vgl. Wooldridge (2002), S. 55 ff.

⁹⁹Wooldridge (2002), S. 50.

¹⁰⁰Vgl. Wooldridge (2002), S. 50f.

¹⁰¹Primärquellen: Durbin (1954), Wu (1974) und Hausman (1978). Bei diesem Test werden die Schätzer der OLS-Regression mit denen einer Instrumentvariablen-Regression (näheres hierzu siehe Abschnitt 5.2.1) verglichen. Die Residuen der Instrumentvariablen-Regression werden hierfür in einer Zusatzregression (nähere Informationen siehe auch Heij et al. (2004)) auf die Residuen der OLS-Regression zuzüglich dessen Regressoren geschätzt. Das sich ergebende R^2 wird ähnlich dem Breusch-Pagan-Test zur Berechnung des Lagrange-Multiplikators verwendet. Er berechnet sich demzufolge aus $LM_{DWH} = nR^2$. Dieser ist näherungsweise $\chi^2(r)$ verteilt, wobei r die Anzahl der möglichen endogenen Regressoren beschreibt.

¹⁰²Hierbei ist zu beachten, dass die Modellspezifikation erst in Abschnitt 5.2.1 beschrieben wird, jedoch die sich darauf aufbauenden P-Werte aus dem Hausmann-Test bereits in Abschnitt 5.1.2 gezeigt werden.

¹⁰³Näheres siehe Backhaus et al. (2006 (11. Auflage)), S. 88 f.

5 Schätzmethoden und Ermittlung der nationalen Bildungsrendite

weise das BIP eines Jahres nicht unabhängig von seiner Ausprägung im Vorjahr. Dieses¹⁰⁴ führt zu ineffizienten Schätzern und demzufolge zu Schwierigkeiten bei der Durchführung von Signifikanztests.

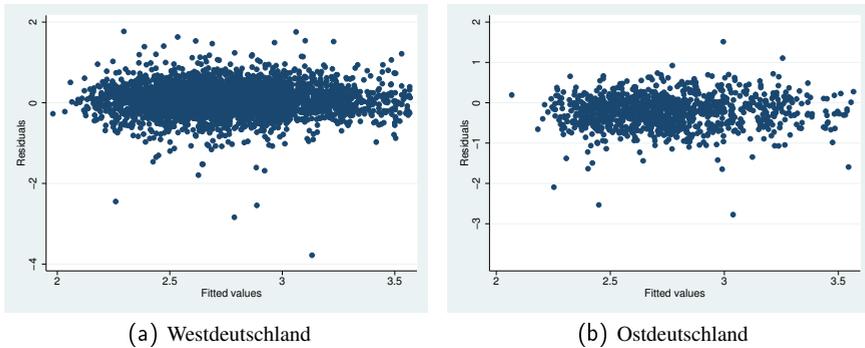


Abbildung 5.2: Optische Prüfung räumlicher Autokorrelation; Quelle: DIW, eigene Berechnungen.

In dieser Arbeit wird eine Querschnittsanalyse vorgenommen. Zeitliche Korrelationen bestehen demnach nicht. Eine weitere Form der Autokorrelation kann sich bei räumlicher Betrachtung ergeben. In diesem relativ neuen Forschungsgebiet werden Korrelationen zumeist auf makroökonomischer Ebene betrachtet. So werden zum Beispiel Zusammenhänge angrenzender Bundesländer berücksichtigt. Auf der mikroökonomischen Ebene wird bisher die räumliche Autokorrelation aufgrund von vielfältiger Heterogenität der Individuen nicht betrachtet. Eine grobe Prüfung kann aber über die Zusammenhänge¹⁰⁵ zwischen vorhergesagten Werten (fitted values) und Residuen erfolgen. Ist ein Muster erkennbar, muss von einer gegebenen Korrelation ausgegangen werden. In den Abbildungen 5.2 sind die Zusammenhänge dargestellt, wobei sich die erste Grafik auf Westdeutschland und die zweite auf Ostdeutschland bezieht. Die Betrachtung für den ostdeutschen Raum zeigt tendenziell negative Residuals. Dies deutet auf eine Überschätzung der ermittelten logarithmierten Bruttostundenlöhne hin. Eine Betrachtung der unterschiedlichen Regionen in Deutschland erweist sich daher als sinnvoll. Diese Betrachtung wird in Abschnitt 6 gezeigt. Für Westdeutschland ist kein Muster erkennbar.

¹⁰⁴Getestet werden kann die zeitliche Autokorrelation über den Durbin-Watson-Test.

¹⁰⁵Siehe Backhaus et al. (2006 (11. Auflage)), S. 87.

V6 Zwischen den unabhängigen Variablen besteht keine lineare Abhängigkeit, d. h. die perfekte Multikollinearität wird ausgeschlossen.

Sind zwei Regressoren perfekt multikollinear, so ist eine der betreffenden Variablen redundant und liefert somit keinen eigenen Erklärungsbeitrag. Jedoch kann auch nichtperfekte Multikollinearität zu Problemen führen. Bei ihrem Vorhandensein können die gleichen Informationen der Variablen zur Schätzung nicht genutzt werden, sie beeinflussen jedoch die Güte¹⁰⁶ der Regression. Zudem können sich bei Hinzufügen bzw. Weglassen von Variablen die Regressionskoeffizienten stark verändern. Zur Prüfung werden die Varianzinflationsfaktoren¹⁰⁷ bestimmt. Grundsätzlich ist das Problem der Multikollinearität bei fast allen empirischen Untersuchungen vorhanden.¹⁰⁸ Zumindest der VIF des interessierenden Merkmals sollte darunter liegen. Die Darstellung der VIFen wird im Anhang A.3.2 mit jeweiligen Verweisen pro Modell dargestellt.

Kriterien zur Beurteilung von Regressionen.

K1 Globale Prüfung der Regressionsfunktion

Die globale Prüfung¹⁰⁹ hat zum Ziel, die Regressionsfunktion und ihren Aussagegehalt in ihrer Gesamtheit zu beurteilen. Damit wird die Frage geklärt, inwieweit die abhängige Variable mit den Regressoren erklärt werden kann. Die Höhe des Erklärungsbeitrags muss im inhaltlichen Zusammenhang bewertet werden. Es wirken vielfache Indikatoren beeinflussend auf den Lohn, die zum Teil schwer quantifizierbar sind. Demzufolge ist üblicherweise¹¹⁰ ein Erklärungsbeitrag über 30 % zufriedenstellend. Für diese Prüfung wird das Bestimmtheitsmaß und die F-Statistik angewendet.

¹⁰⁶Siehe hierzu die Beschreibung des Bestimmtheitsmaß in diesem Abschnitt.

¹⁰⁷Kurz VIF, näheres siehe Backhaus et al. (2006 (11. Auflage)), S. 91 oder Eckey et al. (2004), S. 92. Pro Regressor wird eine Schätzung auf die anderen unabhängigen Variablen durchgeführt, um den etwaigen Erklärungsgehalt eines Regressors durch einen anderen zu bestimmen. Das sich ergebende R^2 pro Zusatzregression wird zur Berechnung des Faktors verwendet. Die Berechnungsformel lautet: $VIF_j = \frac{1}{1-R_j^2}$.

¹⁰⁸Als Faustregel schlägt Eckey et al. (2004, S. 92), einen kritischen Wert von 20 vor.

¹⁰⁹Vgl. Backhaus et al. (2006 (11. Auflage)), S. 63 ff.

¹¹⁰Siehe hierzu die in Abschnitt 3 dargestellten Studien.

Bestimmtheitsmaß

„Das Bestimmtheitsmaß misst die Güte der Anpassung der Regressionsfunktion an die empirischen Daten“¹¹¹. Dieses Maß liegt normiert zwischen 0 und 1, wobei die untere Grenze für keine und die obere Grenze für die komplette Erklärung der Streuung steht. Problematisch ist, dass bei dieser Kennzahl die Menge der Regressoren positiv auf das Bestimmtheitsmaß Einfluss nimmt. Auch nicht nutzenstiftende Regressoren erhöhen das Bestimmtheitsmaß. Um diese verfälschenden Informationen zu entfernen, ist das um die Freiheitsgrade korrigierte Bestimmtheitsmaß¹¹² zu verwenden. Dieses Maß erklärt die Güte der Regressionsfunktion auf Basis der Stichprobe. Das eigentliche Ziel ist es, Aussagen für die Grundgesamtheit zu treffen. Die Übertragbarkeit der Stichprobe auf die Grundgesamtheit muss demnach in einem weiteren Schritt getestet werden.

F-Test

Mit einer empirischen Untersuchung soll ein Sachverhalt aus der Realität bestmöglich geprüft werden. Dafür muss die Stichprobe die Grundgesamtheit widerspiegeln. Hierfür wird die F-Statistik verwendet, die den Stichprobenumfang zur Beurteilung mit einbezieht. Wenn der regressierte kausale Zusammenhang auch auf die Grundgesamtheit übertragbar ist, so müssen die wahren Regressionskoeffizienten β_j von null verschieden sein.¹¹³ Ist die Nullhypothese¹¹⁴ bei gegebener Irrtumswahrscheinlichkeit abzulehnen, so ist davon auszugehen, dass auch in der Grundgesamtheit tatsächlich ein Zusammenhang existiert. Damit ist jedoch nicht klar, ob der einzelne Regressor einen Erklärungsbeitrag liefert.

K2 Spezielle Prüfung der einzelnen Koeffizienten

Ob der einzelne Koeffizient einen signifikanten Beitrag zur Erklärung der abhängigen Variablen liefert, wird über den t-Test¹¹⁵ geprüft. Das Ziel ist es, die Nullhypothese, d. h. die These, dass die Variable keinen Beitrag zur Erklärung der abhängi-

¹¹¹Backhaus et al. (2006 (11. Auflage)), S. 66. Es wird berechnet durch $R^2 = \frac{\text{erklärte Streuung}}{\text{gesamte Streuung}}$.

¹¹²Vgl. Backhaus et al. (2006 (11. Auflage)), S. 68, Berechnungsformel: $R_{\text{kor}}^2 = R^2 - \frac{J(1-R^2)}{n-J-1}$, wobei n: Stichprobenumfang, J: Regressorenzahl.

¹¹³ $H_0 = \beta_1 = \beta_2 \dots = \beta_j = 0$ gilt es abzulehnen.

¹¹⁴Vgl. Backhaus et al. (2006 (11. Auflage)), S. 70 und Heij et al. (2004), S. 164; Berechnungsformel: $F = \frac{n-J}{J-1} \frac{R^2}{1-R^2}$.

¹¹⁵Dabei wird die Nullhypothese: $H_0 : \beta_j = 0$ getestet. Es gilt $t = \frac{\beta_j}{s_{\beta_j}}$, wobei β_j der Regressionskoeffizient des j-ten Regressors und s_{β_j} die zugehörige Standardabweichung darstellt. Vgl. Backhaus et al. (2006 (11. Auflage)) S. 73.

gen Variable liefert, abzulehnen. Wenn das geschieht, kann von einem signifikanten Erklärungsbeitrag des Regressors ausgegangen werden. Die Prüfung wird pro Koeffizient in den Ergebnistabellen für drei Signifikanzniveaus dargestellt.

5.1.1 Modellbeschreibungen

Da der Fokus dieser Arbeit auf der Schätzung einer Bildungsrendite liegt, wird das Hauptaugenmerk auf den Koeffizienten der Bildungsdauer gelegt. Die weiteren Variablen werden als Kontrollvariablen mit hinzugezogen, um einen besseren Erklärungsgehalt zu erreichen sowie um Faktoren, die sowohl die Bildungsrendite als auch das Einkommen beeinflussen, herauszurechnen. Jedoch können nicht alle Einflüsse aufgrund mangelnder Quantifizierbarkeit in der Gleichung berücksichtigt werden. In Abschnitt 5.1 in Voraussetzung V1 wird dies diskutiert.

Bereits in den Abschnitten 4.2 und 4.3 wird näher auf die Modellvariablen eingegangen. Der Zusammenhang der erklärenden Variablen zum Einkommen wurde dort motiviert und über deskriptive Statistiken und eine bivariate Analyse geprüft. Die Modelldarstellungen in diesem Abschnitt werden daher hier nur zusammenfassend dargestellt. Die Bezeichnung, Skalierung und der vermutete Zusammenhang zum Bruttostundenlohn der jeweiligen Variablen werden kurz gezeigt, nicht nochmals motiviert.

Modell 1 – Grundmodell nach J. Mincer. Die Bildungsrendite zeigt den Einfluss der Bildungsdauer auf den logarithmierten Bruttostundenlohn. In der Schätzgleichung wird sie als Koeffizient der Bildungsdauer, der sich aus der Lohngleichung ergibt, ausgedrückt. Dieser Schätzer stellt den semilogarithmischen Zusammenhang dar und muss, um die prozentuale Rendite zu ermitteln, zur Interpretation transformiert werden. Das Basismodell von Jacob Mincer¹¹⁶ wird in Abschnitt 2 dargestellt und bildet das Gerüst für die Schätzgleichung.¹¹⁷ Demzufolge wird im ersten Modell folgende Lohngleichung spezifiziert:

$$\ln Y_i = \beta_0 + \beta_1 S_i + \beta_2 Erf_i + \beta_3 Erf_i^2 + \varepsilon_i.$$

Dabei ist die zu erklärende Variable $\ln Y_i$ der logarithmierte Bruttostundenlohn des Individuums i . Die Koeffizienten sind durch β_j angegeben und sind ein Maß für die Stärke

¹¹⁶Siehe hierzu Mincer (1974).

¹¹⁷Die Koeffizienten werden in den verschiedenen Modellen stets mit β_j bezeichnet. Sie beziehen sich jeweils auf das gerade betrachtete Modell.

des Zusammenhangs zum Einkommen. Der Koeffizient β_1 zeigt die Bildungsrendite und wurde in Abschnitt 2 mit r abgekürzt. Er zeigt die prozentuale¹¹⁸ Verzinsung eines zusätzlichen Bildungsjahres. Die weiteren Variablenbezeichnungen sind in Tabelle 5.1 näher erläutert. Ebenso ist die Skalierung des Merkmals sowie der vermutete Zusammenhang zum Bruttostundenlohn dargestellt.

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
S_i	Bildungsdauer	metrisch	positiv
Erf_i	Erfahrung	metrisch	positiv, im Quadrat negativ

Tabelle 5.1: Variablenbezeichnung – Grundmodell.

Die nähere Beschreibung der Variablen samt deskriptiver Statistiken findet sich in Abschnitt 4.2.1. Ausgehend von der Humankapitaltheorie nach Gary Becker¹¹⁹ ist von einem positiven Zusammenhang zwischen der Bildungsdauer und dem Einkommen auszugehen. Gleiches gilt für die Berufserfahrung. Bei dieser ist der mit zunehmender Erfahrung sinkende Grenzertrag zu berücksichtigen. Daher wird die Erfahrung in ihrem Quadrat zusätzlich implementiert. Der Theorie folgend, wird für diesen Term wird von einem negativen Einfluss auf das Einkommen ausgegangen, demzufolge gilt $\beta_3 < 0$.

Modell 2 – Persönliche Gegebenheiten. Der Einfluss von persönlichen Gegebenheiten auf das Einkommen ist groß. Das Modell wird deshalb um drei Variablen erweitert. Zum einen wird das Geschlecht einbezogen, um damit das üblicherweise geringere Einkommen von Frauen zu berücksichtigen. Demzufolge ist von einem negativen Koeffizienten auszugehen. Mit der Berücksichtigung des Familienstandes wird der positive Einfluss der sozialen Vernetzung hinzugezogen. Daher wird vermutet, dass die Dummy-Variable ein positives Vorzeichen hat. Im Gegenzug dazu jedoch muss für verheiratete Frauen von einem negativen Einfluss auf das Einkommen ausgegangen werden. Dieser Effekt wird über einen Interaktionsterm in der Gleichung berücksichtigt. Dabei nimmt diese Variable den Wert 1 an, wenn das Individuum eine verheiratete Frau ist, in den anderen Fällen wird die Ausprägung null zugeordnet.

¹¹⁸Aufgrund des semilogarithmischen Zusammenhangs ist zuvor eine Rückrechnung notwendig.

¹¹⁹Siehe hierzu Becker (1994 (3. Auflage)).

5.1 Schätzverfahren – Methode der kleinsten Quadrate (OLS)

Im Folgenden wird auf das bereits diskutierte Modell aufgebaut. Daher sollen die zuvor beschriebenen Kontrollvariablen über einen Summenterm¹²⁰ verkürzt dargestellt werden. Einzige Ausnahme ist die Variable der Bildungsdauer, um deren Bedeutung für die vorliegende Arbeit hervorzuheben. Das neue Modell lautet demnach

$$\ln Y_i = \beta_0 + \beta_1 S_i + \sum_{j=2}^3 \beta_j X_{ij} + \beta_4 \text{weibl}_i + \beta_5 \text{verhei}_i + \beta_6 \text{IAT}_{w-\text{verh}_i} + \varepsilon_i.$$

Die Variablenbezeichnungen sind in Tabelle 5.2 zusammengefasst.

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
<i>weibl_i</i>	weiblich	dichotom (1: weiblich, 0: männlich)	negativ
<i>verhei_i</i>	Familienstand	dichotom (1: verheiratet, zusammenlebend, 0: sonst)	positiv
<i>IAT_{w-verh_i}</i>	Interaktions-term	dichotom (1: verheiratet, zusammenlebend <i>und</i> weiblich, 0: sonst)	negativ

Tabelle 5.2: Variablenbezeichnung – Persönliche Gegebenheiten.

Modell 3 – Unternehmensgröße. Die genaue Beschreibung der Dummy-Variablen der Unternehmensgröße findet sich in Abschnitt 4.2.1. Es ist davon auszugehen, dass die Größe des Unternehmens positiv auf das absolute individuelle Einkommen wirkt. So wird das aktuelle Modell um eine Dummy-Variable erweitert. Das neue Modell lautet nun

$$\ln Y_i = \beta_0 + \beta_1 S_i + \sum_{j=2}^{n=6} \beta_j X_{ij} + \beta_7 D_{\text{groß}_i} + \varepsilon_i.$$

Die Variablenbeschreibung befindet sich in Tabelle 5.3 und wird zusammen mit der Skalierung und dem vermuteten Zusammenhang zwischen Unternehmensgröße und Bruttostundenlohn dargestellt.

¹²⁰Dieser wird pro Modellerweiterung zunehmend ergänzt. Die Kontrollvariablen werden verkürzt dargestellt mit X_{ij} .

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
$D_{gro\beta_i}$	großes Unternehmen	dichotom (1: mehr als 200 Mitarbeiter im Unternehmen, 0: sonst)	positiv

Tabelle 5.3: Variablenbezeichnung – Unternehmensgröße.

Modell 4 – Wohnort. Die Region ist ein einkommensbeeinflussender Faktor. Allein die Unterscheidung, ob das betrachtete Individuum in einem urbanen oder eher ländlichem Gebiet lebt, hat einen Einfluss auf das Einkommen. Dabei wird zumeist davon ausgegangen, dass jene in ländlichen Gebieten ein geringeres Einkommen haben. Weiterhin ist der Wohnort eines Individuums nicht zwangsläufig gleich dem Arbeitsort. Häufig spielen Pendelbewegungen eine entscheidende Rolle. Einige Berufe setzen diese Flexibilität zwingend voraus, andere Arbeitnehmer nehmen freiwillig eine Distanz zwischen Wohn- und Arbeitsort in Kauf. Zu vermuten ist, dass Pendler ein höheres Einkommen haben. Das neue Modell wird erweitert auf:

$$\ln Y_i = \beta_0 + \beta_1 S_i + \sum_{j=2}^{n=7} \beta_j X_{ij} + \beta_8 D_{l\ddot{a}ndl_i} + \beta_9 D_{P\ddot{e}ndler_i} + \varepsilon_i.$$

Die Variablenbeschreibung findet sich in Tabelle 5.4, ebenso wie die Skalierung und der vermutete Zusammenhang zum Bruttostundenlohn.

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
$D_{l\ddot{a}ndl_i}$	ländliches Gebiet	dichotom (1: Wohnort ist mehr als 25 km vom Stadtzentrum entfernt, 0: sonst)	negativ
$D_{P\ddot{e}ndler_i}$	Pendler	dichotom (1: mindestens wöchentliches Pendeln, 0: sonst)	positiv

Tabelle 5.4: Variablenbezeichnung – Wohnort.

5.1.2 Bildungsrenditen in Deutschland

In Tabelle 5.5 werden die Ergebnisse der in Abschnitt 5.1.1 beschriebenen Modelle dargestellt. Die Regression von Modell 1 ergibt eine Bildungsrendite in Höhe von 8,9 %. Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs stets wie folgt berechnet:

$$\text{prozentuale Bildungsrendite} = (e^{\beta_1} - 1) \cdot 100.$$

Interpretiert werden kann dieser Anteil als die mittlere Zunahme des Einkommens um 8,9 % bei Generierung eines zusätzlichen Bildungsjahres. Vergleicht man die Bildungsrenditen bei Hinzunahme weiterer Kontrollvariablen, so können nur leichte Veränderungen in ihrer Höhe festgestellt werden. Sie ist dabei stets, als eine Verzinsung pro zusätzlichem Bildungsjahr zu verstehen. Die durchschnittliche Verzinsung von Humankapital liegt dabei zwischen 8,5 % und 9,0 %. Dieser Koeffizient für die Bildungsrendite erweist sich als stabil. Für alle Modellspezifikationen ist der Einfluss der Bildungsdauer auf das Einkommen signifikant von null verschieden.

Der vermutete Zusammenhang zwischen Berufserfahrung und logarithmiertem Bruttostundenlohn kann ebenfalls bestätigt werden. So wirkt zunehmende Berufserfahrung positiv auf das Einkommen. Auch diese Koeffizienten unterscheiden sich je nach Modell nur geringfügig. Für den linearen Einfluss der Berufserfahrung kann eine Rendite zwischen 2,9 % und 3,3 % geschätzt werden. Der quadrierte Term der Berufserfahrung zeigt für alle Modelle einen signifikant negativen Koeffizienten. Dies bedeutet, dass mit zunehmender Berufserfahrung der Anstieg des Einkommens weniger stark ist als für einen Berufseinsteiger.

Bei der Berücksichtigung der persönlichen Gegebenheiten zeigt sich, dass das Einkommen einer Arbeitnehmerin geringer ist. Der Koeffizient schwankt hierbei zwischen 13,2 % und 14,8 %, d. h. eine Frau verdient ca. 14 % weniger als ein gleichqualifizierter Mann. Fraglich dabei ist, ob sich dies durch eine Selbstselektion der Frauen und Männer auf bestimmte Tätigkeiten ergibt. Der Einfluss des Familienstandes ergibt, dass eine Ehe durch die Nutzung von Spezialisierungsgewinnen eine Einkommenssteigerung zwischen 12,2 % und 13,4 % bewirkt.

Um den Effekt für eine verheiratete Frau – gemessen über den Interaktionsterm – zu interpretieren, müssen alle Variablen der persönlichen Gegebenheiten berücksichtigt werden. Hierfür werden die einzelnen Schätzer für das Geschlecht und den Familienstand sowie

5 Schätzmethoden und Ermittlung der nationalen Bildungsrendite

der Koeffizient des Interaktionsterms aufaddiert. Daraus ergibt sich für eine veheiratete¹²¹ Frau eine Einkommensminderung zwischen 16,5 % und 19,6 %. Damit müssen verheiratete Arbeitnehmerinnen noch deutlichere Einkommenseinbußen hinnehmen als es bei unverheirateten Frauen der Fall ist.

Variable	Modell 1	Modell 2	Modell 3	Modell 4
Bildungsjahre	0,085*** [0,002]	0,086*** [0,002]	0,083*** [0,002]	0,082*** [0,002]
Erfahrung	0,032*** [0,004]	0,030*** [0,004]	0,029*** [0,004]	0,029*** [0,004]
Erfahrung ²	-0,000*** [0,000]	-0,000*** [0,000]	-0,000*** [0,000]	-0,000*** [0,000]
weiblich		-0,138*** [0,022]	-0,127*** [0,021]	-0,124*** [0,021]
verheiratet		0,126*** [0,019]	0,120*** [0,018]	0,115*** [0,018]
IAT verh. Frau		-0,167*** [0,026]	-0,150*** [0,025]	-0,144*** [0,025]
Unternehmen mit mehr als 200 Mitarbeitern			0,226*** [0,012]	0,221*** [0,012]
ländliches Gebiet				-0,053*** [0,012]
Pendler				0,063*** [0,012]
Konstante	1,149*** [0,060]	1,341*** [0,065]	1,268*** [0,063]	1,259*** [0,065]
korr. R ²	0,21	0,29	0,34	0,35
F-Test	0,000	0,000	0,000	0,000
JB-Test	0,000	0,000	0,000	0,000
BP-Test	0,000	0,000	0,000	0,000
VIF _S	1,1	1,1	1,1	1,1
DWH-Test	0,000	0,000	0,000	0,000
n	4.693	4.693	4.693	4.693

Tabelle 5.5: Ergebnistabelle der OLS-Regression; in eckigen Klammern sind die Standardfehler nach White mit den Signifikanzniveaus: *** 1 %, ** 5 % und * 10 % angegeben. Die Testergebnisse werden über die P-Werte gezeigt. Zur Anwendung des DWH-Tests wird das 4. Modell der IV-Schätzung als Vergleich verwendet. Quelle: DIW, eigene Berechnungen.

¹²¹ Auch Frauen, die in einer eheähnlichen Gemeinschaft leben, zählen dazu.

Der Einfluss der Unternehmensgröße auf das Einkommen ist groß. Aus der Schätzung resultiert, dass zwei gleich qualifizierte Arbeitnehmer einen um ca. 25 % verschiedenen Bruttostundenlohn, nur aufgrund der Unternehmensgröße erhalten. Dies könnte auf produktivere Arbeitnehmer zurückzuführen sein.

Modell 4 zeigt die Erweiterung um die wohnortspezifischen Variablen. Hierbei ergibt sich ein negativer Koeffizient für Personen, die im ländlichen Raum wohnen. Damit beziehen Individuen, die mehr als 25 km vom nächsten Stadtzentrum entfernt wohnen, im Mittel ein um 5,4 % geringeres Einkommen. Betrachtet man dagegen Pendler, so zeigt sich eine einkommenssteigernde Wirkung. Dies bedeutet, dass Arbeitnehmer, die zumindest ein Mal wöchentlich pendeln, im Mittel 6,5 % mehr verdienen als jene, die an ihrem Arbeitsort leben.

Die Betrachtung der Inferenzstatistik zeigt, dass alle Koeffizienten einen signifikanten Beitrag zur Erklärung des logarithmierten Bruttostundenlohns liefern. Das korrigierte R^2 kann durch die Modellerweiterungen beständig erhöht werden und liegt im Modell 4 bei einem Anteil von 35 % Erklärungsgehalt der Streuung an der Gesamtstreuung. Auch der F-Test bestätigt für alle Modelle den generellen Zusammenhang zwischen den erklärenden und der erklärten Variablen. Der Test auf normalverteilte Residuen von Jarque und Bera muss für alle Modellspezifikationen abgelehnt werden. Doch aufgrund der großen Beobachtungszahlen kann approximativ von normalverteilten Koeffizienten ausgegangen werden, sodass statistische Tests trotzdem näherungsweise anwendbar sind. Der Breusch-Pagan-Test auf Homoskedastizität ist ebenfalls abzulehnen. Eine Korrektur der Standardfehler erfolgt über das in Abschnitt 5.1 beschriebene Verfahren von White. Die korrigierten Fehler sind pro Koeffizient in eckigen Klammern angegeben. Die Prüfung der Multikollinearität erfolgt über die Varianzinflationsfaktoren (VIF). Da pro Regressor ein VIF zu berechnen ist, werden diese im Anhang A.3.2 in Tabelle A.11 angegeben. Einzig der VIF für die Variable der Bildungsdauer wird zusätzlich noch in Tabelle 5.5 gezeigt. Er hat für alle Modellspezifikationen den Wert 1,1. Da dieser deutlich unterhalb des kritischen Wertes liegt, kann davon ausgegangen werden, dass die Bildungsdauer nur marginal mit den weiteren erklärenden Variablen korreliert ist. Das Multikollinearitätsproblem ist somit als gering einzustufen. Auch für die Kontrollvariablen kann nicht von einem gravierenden Multikollinearitätsproblem ausgegangen werden.¹²² Einzig die Erfahrungsvariable bildet hier eine Ausnahme, da sie sowohl in linearer als auch quadratischer Form implementiert ist. Für den Test auf Exogenität der Regressoren ist ein Vergleich mit einer Instrumentvariablen-

¹²²Siehe hierzu A.3.2.

Regression¹²³ nötig. Das Referenzmodell ist in Abschnitt 5.2.1 näher beschrieben. Die hier in Tabelle 5.5 dargestellten Ergebnisse basieren auf Modell 4 der IV-Regression. Für alle beschriebenen Modellspezifikationen muss die Exogenitätsannahme abgelehnt werden. Dies deutet auch mögliche inkonsistente Schätzer bei Nutzung der OLS-Methode hin. Demnach ist davon auszugehen, dass die sich ergebenden Schätzer verzerrt sind und die Anwendung anderer Verfahren in Betracht gezogen werden muss.

Bereits in Abschnitt 5.1 sind einige Voraussetzungen des OLS-Verfahrens beschrieben und geprüft worden. Zusätzlich sind in Tabelle 5.5 die P-Werte pro Modell dargestellt. Aus diesen Prüfungen ergibt sich, dass die Voraussetzungen V1 bis V6 mit Ausnahme von V4 zumindest näherungsweise als gegeben angenommen, korrigiert oder als ein nur marginal beeinflussendes Problem bewertet werden können. Nach Vergleich aller Modelle zeichnet sich das 4. mit der besten Güte aus. Da für die anderen Spezifikationen die Eigenschaften im Ergebnis übereinstimmen, wird dieses als Referenzmodell für weitere Betrachtungen verwendet. Im besonderen Fokus muss dabei zunächst die Behebung des Endogenitätsproblems (V4) liegen. Dies erfolgt über die Anwendung weiterer Methoden zur Schätzung der Bildungsrendite.

5.1.3 Zusammenfassung

In diesem Abschnitt werden verschiedene Modellspezifikationen mit dem OLS-Verfahren regressiert. Ziel dabei ist es, das bestmögliche Modell zur Schätzung der Bildungsrendite zu identifizieren. Dabei werden die dargestellten notwendigen Voraussetzungen für die Durchführung dieses Verfahrens geprüft. Es ergibt sich, dass alle Spezifikationen die gleichen Stärken und Schwächen hinsichtlich der Erfüllung der Voraussetzungen aufweisen. Demzufolge wird jenes Modell ausgewählt, welches die beste Güte aufweist. Mit dieser Spezifikation, Modell 4, wird eine mittlere jährliche Verzinsung von Bildung in Höhe von 8,5 % geschätzt. Nachweislich ist die Variable der Bildungsdauer endogen und erfüllt demnach nicht die Voraussetzung der Exogenität zur Ermittlung von unverzerrten Schätzern. Alle anderen Voraussetzungen sind entweder erfüllt, können bereinigt werden oder haben keinen Einfluss auf den Koeffizienten zur Schätzung der Bildungsrendite, der im Fokus dieser Arbeit steht. Im Weiteren gilt es daher, das Endogenitätsproblem zu lösen. Hierfür werden zusätzliche Verfahren hinzugezogen.

¹²³Nachfolgend mit IV-Regression oder 2SLS-Verfahren abgekürzt.

5.2 Mehrgleichungsmodelle zur Schätzung von Bildungsrenditen

Wie in Abschnitt 5.1.2 dargestellt, muss davon ausgegangen werden, dass die Voraussetzung der Exogenität der Regressoren (V4) nicht erfüllt ist. Um dem Problem zu begegnen, werden Mehrgleichungsmodelle angewendet. Dabei gilt als Basis das Modell 4 aus Abschnitt 5.1.2. Die Lohngleichung dieser Spezifikation lautet:

$$\begin{aligned} \ln Y_i &= \beta_0 + \beta_1 S_i + \beta_2 Erf_i + \beta_3 Erf_i^2 + \beta_4 weibl_i + \beta_5 verhei_i \\ &+ \beta_6 IAT_{w-verh_i} + \beta_7 D_{gro\beta_i} + \beta_8 D_{l\ddot{a}ndl_i} + \beta_9 D_{Pendler_i} + \varepsilon_i. \end{aligned}$$

Das identifizierte Problem der Endogenität kann verschiedene Ursachen haben. Es entsteht durch

- das Fehlen von relevanten Variablen („*omitted variables*“): Häufig gibt es Einflussgrößen, die nicht identifizierbar bzw. nicht quantifizierbar sind, jedoch einen wichtigen Beitrag zur Erklärung der abhängigen Variablen liefern.
- Selbstselektion („*sample selection*“): Die Selbstselektion einzelner Gruppen in verschiedene Teilgruppen ist schwerlich identifizierbar, verletzt jedoch die Heterogenitätsbedingung.
- Messfehler im Regressor: Ein systematischer Messfehler einer oder mehrerer erklärender Variablen kann zu systematischen Verzerrungen führen.

Mögliche Messfehler müssen bei Vorhandensein hingenommen werden. Für den hier verwendeten Datensatz kann durch kritische Prüfung vermutet werden, dass sich die Messfehler und ihre Auswirkungen in Grenzen halten, auch wenn sie nicht gänzlich auszuschließen sind. Das Problem der *omitted variables* sowie der *sample selection* wird nachfolgend ausführlich betrachtet. Dabei wird zunächst das erste Problem untersucht. Die Selektionsproblematik wird in Abschnitt 5.4.3 ausführlich diskutiert.

5.2.1 Two-Stage Least Squares

Liegt ein Endogenitätsproblem¹²⁴ aufgrund von einer nicht meßbaren, daher fehlenden Variable vor, die zudem multikollinear mit der endogenen Variablen verknüpft ist, so ist diese mit dem Fehlerterm korreliert. Dies bedeutet, dass es einen Einfluss einer dritten Größe gibt, die auf einen oder mehrere Regressoren wirkt. Im Ergebnis dieser mangelnden Identifizierbarkeit und Quantifizierbarkeit ist der OLS-Schätzer

- verzerrt.
- inkonsistent. Damit wird *„der Bias auch bei großem Umfang der verfügbaren Daten nur dann vernachlässigbar sein [...], wenn die Korrelation zwischen den Regressoren und den Störgrößen asymptotisch gegen null geht.“*¹²⁵
- ineffizient. Demnach ist die Anwendung von Teststatistiken ungültig.

Zusammenfassend sind bei einem gegebenen Endogenitätsproblem durch fehlende Variablen die Ergebnisse der OLS-Regression wenig vertrauenswürdig. Der Nachweis erfolgt über den Test von Durbin, Wu und Hausman. Das Problem kann über die Methode der Momente¹²⁶ bereinigt werden. Im Speziellen wird hier die Instrumentvariablenschätzung (IV-Schätzung), genauer die 2SLS-Methode verwendet.¹²⁷ Ausgangspunkt ist dabei die vorhandene Korrelation zwischen zumindest einem Regressor und dem Fehlerterm. Es gilt demnach $E(x_{ij}, \varepsilon_i) \neq 0$. Ziel ist es, zwischen Regressor und Fehlerterm Unabhängigkeit herzustellen. Dafür werden Instrumentvariablen z_i' ermittelt, für die gilt, dass sie mit dem endogenen Regressor korrelieren, aber unkorreliert sind mit dem Residuum der Ausgangsregression.

Der IV-Ansatz ist ein zweistufiges sequentielles Verfahren. Zunächst wird der endogene Regressor als neue abhängige Variable auf ausgewählte Instrumente und die ursprünglich weiteren exogenen Variablen geschätzt. Diese Regression wird als Reduced-Form-Regression¹²⁸ bezeichnet. Die geschätzten Werte für den Regressand können in einem zweiten Schritt anstelle der nachweislich endogenen Variable implementiert werden. Wurde die Variable mit sinnvollen¹²⁹ Instrumenten geschätzt, kann nun von der Exogenität dieser ausgegangen werden. Die besondere Schwierigkeit bei der Verwendung des IV-

¹²⁴Nähere Erläuterungen finden sich in Voraussetzung 4 in Abschnitt 5.1.

¹²⁵Hackl (2007), S. 255.

¹²⁶Siehe hierzu Mátyás (1999).

¹²⁷Künftig werden die Begriffe Instrumentvariablen-Schätzung und die 2SLS-Methode synonym verwandt.

¹²⁸Oder auch First-Stage-Regression.

¹²⁹Die Diskussion dieser erfolgt im weiteren Verlauf dieses Abschnitts.

Ansatzes liegt in der Auswahl geeigneter Instrumente. Es ist eine Herausforderung, Instrumente zu finden, die einerseits die endogene Variable erklären und andererseits keine Korrelation mit dem Fehlerterm der Ausgangsregression aufweisen. Die Voraussetzungen für diese Methode werden zusammen mit der jeweiligen Modellüberprüfung nach den Modellbeschreibungen detailliert diskutiert.

Bezug zur Bildungsrendite. Wie bereits in Abschnitt 5.1.2 durch den Test von Durbin, Wu und Hausmann¹³⁰ gezeigt, muss von einem Endogenitätsproblem für die Variable der Bildungsdauer in der Lohngleichung ausgegangen werden. Dies bedeutet, dass es eine sowohl das Einkommen als auch die Bildungsdauer beeinflussende Größe¹³¹ gibt, die zur Verzerrung des Schätzers führt. Dies wird in der Literatur als das Ability-Bias-Problem¹³² bezeichnet. Beschrieben werden damit die meist unbeobachtbaren Charaktereigenschaften, wie individuelle Fähigkeiten und Motivationen eines Arbeitnehmers. Ziel ist es, den reinen Effekt der Bildung auf das Einkommen zu ermitteln, unabhängig vom IQ oder Ähnlichem. „Wird dafür bei der Schätzung der Lohnfunktion nicht kontrolliert, gibt der geschätzte Koeffizient der Bildungsvariablen nicht den kausalen Effekt eines zusätzlichen Schuljahres auf den Lohn [...] an.“¹³³

Abbildung 5.3.a veranschaulicht das Problem, wobei F_i die Begabung bzw. die Fähigkeiten beschreibt. Diese individuell verschiedenen Motivationen wirken sowohl auf das Bildungsniveau, als auch auf das Einkommensniveau. Das Endogenitätsproblem ergibt sich aufgrund der mangelnden Messbarkeit dieser Variablen. Häufig muss daher die Fähigkeiten-Variable in der Analyse unberücksichtigt gelassen werden. Dies führt zu einer Korrelation zwischen der Bildungsdauer und dem Fehlerterm. Die Bildungsdauer ist demnach nicht mehr exogen. Dargestellt ist dies in Abbildung 5.3.b. Die Folge ist ein verzerrter, ineffizienter und inkonsistenter Schätzer für die Bildungsrendite. Um dem zu begegnen wird in vielen Studien die Handhabung mit den heterogenen Fähigkeiten diskutiert. Eine Möglichkeit ist es, Schätzwerte für die individuellen Fähigkeiten zu ermitteln und in die Analyse einzubeziehen. Dabei werden häufig Intelligenztests zu Hilfe genommen.

- Die Testergebnisse¹³⁴ werden als Schätzgröße für die Auffassungsgabe und intrinsische Motivation in die Regression einbezogen. Die neue Regressionsgleichung

¹³⁰Siehe hierzu auch Heij et al. (2004), S. 409 ff. oder Hackl (2007), S. 259 f.

¹³¹Siehe hierzu: Griliches (1977) sowie Card (1999).

¹³²Siehe hierzu z. B. Belzil und Hansen (2002).

¹³³Steiner und Lauer (2000).

¹³⁴Siehe hierzu Card (1995).

5 Schätzmethoden und Ermittlung der nationalen Bildungsrendite

würde lauten: $\ln Y_i = \beta_0 + \beta_1 S_i + \beta_2 F_i + \sum_{j>2}^n \beta_j X_{ij} + \varepsilon_i$, wobei F_i die individuellen IQ-Testergebnisse symbolisiert. Diese Ergebnisse stehen näherungsweise für die individuellen Fähigkeiten. Im gegebenen SOEP-Datensatz steht solch eine Variable nicht zur Verfügung. Demnach kann diese nicht in das Modell einbezogen werden.

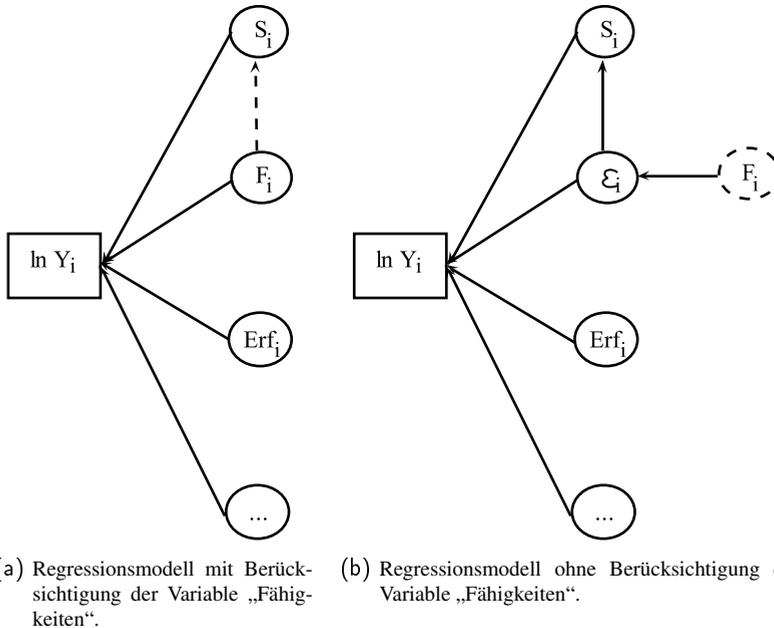


Abbildung 5.3: Der Einfluss der Berücksichtigung der Variable „Fähigkeiten“; eigene Darstellung.

Eine andere Möglichkeit ist die näherungsweise Eliminierung der individuellen Heterogenität. Dabei werden weitestgehend homogene Individuen miteinander verglichen. Es bieten sich hierbei

- Zwillingsstudien an. So durchgeführt zum Beispiel von Ashenfelter und Krueger (1994) mit Daten der Current Population Survey und für Deutschland von Schnabel und Schnabel (2002). Bei Anwendung dieses Verfahrens wird davon ausgegangen, dass Zwillinge von der Erziehung und den genetischen Anlagen eine deutlich größere Homogenität aufweisen und damit hauptsächlich der Bildungsunterschied einkommensentscheidend ist. Diese Möglichkeit ist mit dem gegebenen Datensatz aufgrund zu geringer Fallzahlen nicht umsetzbar.

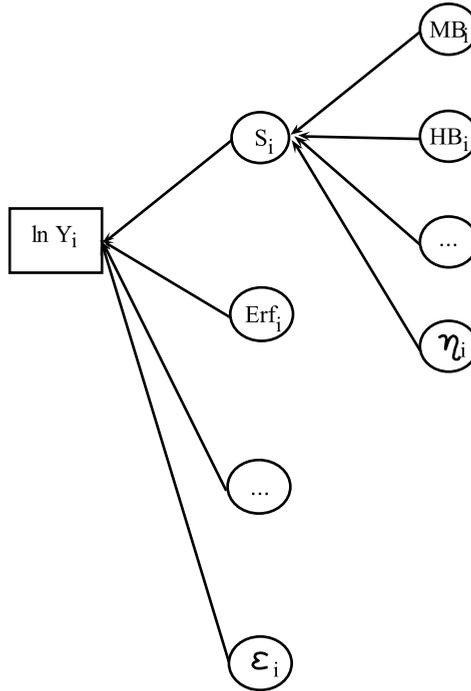


Abbildung 5.4: IV-Verfahren; eigene Darstellung.

Beim dritten Weg wird die Bildungsdauer um den Motivationseinfluss bereinigt. Hierfür wird das zweistufige IV-Verfahren verwendet. Für die hier anstehende erste Regression bedeutet dies, dass „zusätzliche Variablen (Instrumente) in die Analyse mit aufgenommen [werden], die das Bildungsniveau des Einzelnen mitbestimmen, die aber nicht von seinen unbeobachtbaren angeborenen Fähigkeiten abhängig sein dürfen.“¹³⁵ Über die IV-Methode wird demnach angestrebt, eine Bildungsdauer zu schätzen, die unabhängig von den Fähigkeiten und damit vom Fehlerterm ist. Die zweite Regression hat dann das Ziel, dass „ihre 'theoretischen' Werte [...] ersetzt“¹³⁶ und die Lohngleichung damit neu geschätzt wird. Das zweistufige Verfahren wird in Abbildung 5.4 veranschaulicht. Dabei stehen beispielhaft für mögliche Instrumente die Variablen MB_i und HB_i für das mittlere bzw. hohe Bildungsniveau des Vaters. Die detaillierte Diskussion möglicher Instrumente folgt im nächsten Schritt.

¹³⁵Suter (2005), S. 2.

¹³⁶Moosmueller (2004), S. 185.

Modellbeschreibung der Reduced-Form-Regression. Wie aus Abschnitt 5.2.1 ersichtlich, sind die Anforderungen an die Instrumente dieser Regression hoch. In diesem Abschnitt werden daher verschiedene Modellspezifikationen zur Schätzung der Bildungsdauer diskutiert. In Abschnitt 4.2.2 wurden die verwendeten Variablen motiviert sowie durch eine erste deskriptive Analyse veranschaulicht. Abschnitt 4.3.2 zeigte in einer bivariaten Analyse die Korrelationskoeffizienten zur Bildungsdauer. An dieser Stelle werden die verschiedenen Modelle kurz beschrieben sowie der jeweilig vermutete Zusammenhang wird aufgezeigt.

IV-Modell 1 – Elterliche Bildung. In Abschnitt 4.2.2 ist bereits dargestellt, dass lediglich die Bildung des Vaters als soziale Prägung einbezogen werden kann. Dafür werden zwei Dummy-Variablen berücksichtigt, deren Koeffizienten relativ zum niedrigsten Bildungsniveau zu interpretieren sind. Dieses erste Modell zur Schätzung der Bildungsdauer hat die Spezifikation:

$$S_i = \gamma_0 + \gamma_1 D_{MB_i} + \gamma_2 D_{HB_i} + \sum_{j>2}^9 \gamma_j X_{ij} + \eta_i,$$

wobei die zu erklärende Variable diesmal die Bildungsdauer, S_i , ist. Die Kontrollvariablen aus Modell 4¹³⁷ der Lohngleichung, dargestellt in Abschnitt 5.1.1, gehen in die Regression ein und sind innerhalb der Summe zusammengefasst, wobei j der Laufindex der verschiedenen Kontrollvariablen ist. Die weiteren Variablen sind in Tabelle 5.6 beschrieben.

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
D_{MB_i}	mittlere Bildung des Vaters	dichotom (1: mittlerer Bildungsabschluss, 0: sonst)	positiv (in Relation zum niedrigen Bildungsabschluss)
D_{HB_i}	hohe Bildung des Vaters	dichotom (1: hoher Bildungsabschluss, 0: sonst)	positiv (in Relation zum niedrigen Bildungsabschluss)

Tabelle 5.6: IV-Verfahren – Modell 1.

¹³⁷Siehe Abschnitt 5.1.1.

Der besondere Einfluss der sozialen Prägung wird in diversen Studien betont. Auch die deskriptiven Statistiken zeigen diesen Zusammenhang und lassen einen positiven Einfluss auf die Bildungsdauer vermuten.

IV-Modell 2 – Anzahl der Bücher. Bücher stehen als ein Symbol für Bildung. Der intelligente Mensch wird häufig mit einem belesenen gleichgesetzt. Pauschal kann damit vermutet werden, dass die Bildungsdauer durch die Zahl der Bücher im eigenen Haushalt positiv beeinflusst wird. Die Erweiterung des Modells um diese Dummy-Variable führt zu folgendem Modell:

$$S_i = \gamma_0 + \gamma_1 D_{MB_i} + \gamma_2 D_{HB_i} + \gamma_3 N Buecher_i + \sum_{j>3}^{10} \gamma_j X_{ij} + \eta_i,$$

wobei die Variablenbeschreibung mit Skalierung in Tabelle 5.7 zu finden ist.

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
$N Buecher_i$	Anzahl der Bücher im Haushalt	dichotom (1: mehr als 200 Bücher, 0: sonst)	positiv

Tabelle 5.7: IV-Verfahren – Modell 2.

IV-Modell 3 – Anzahl der Geschwister. Die Theorie von Willis (1973) lässt einen umgekehrt proportionalen Zusammenhang zwischen der Geschwisterzahl und der Bildungsdauer vermuten. Bei Hinzunahme der Zahl der Brüder und Schwestern ergänzt sich das Modell zu:

$$S_i = \gamma_0 + \gamma_1 D_{MB_i} + \gamma_2 D_{HB_i} + \gamma_3 N Buecher_i + \gamma_4 N Schw_i + \gamma_5 N Brued_i + \sum_{j>5}^{13} \gamma_j X_{ij} + \eta_i,$$

wobei die Variablenbeschreibung aus Tabelle 5.8 ersichtlich wird.

5 Schätzmethoden und Ermittlung der nationalen Bildungsrendite

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
$N Schw_i$	Anzahl der Schwestern	metrisch	negativ
$N Brued_i$	Anzahl der Brüder	metrisch	negativ

Tabelle 5.8: IV-Verfahren – Modell 3.

IV-Modell 4 – Staatsangehörigkeit. Personen mit nicht deutschem Hintergrund fällt es aufgrund von Sprachbarrieren häufig schwerer eine hohe Bildungsdauer zu erreichen. Daher wird das Modell ergänzt um eine Dummy-Variable für die ausländischen Staatsangehörigkeit und lautet:

$$S_i = \gamma_0 + \gamma_1 D_{MB_i} + \gamma_2 D_{HB_i} + \gamma_3 N Buecher_i + \gamma_4 N Schw_i + \gamma_5 N Brued_i + \gamma_6 D_{Nicht-Deutsch_i} + \sum_{j>6}^{14} \gamma_j X_{ij} + \eta_i,$$

wobei die Variablenbeschreibung in Tabelle 5.9 dargestellt ist.

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
$D_{Nicht-Deutsch}$	Nationalität nicht Deutsch	dichotom (1: sonst, 0: Nationalität: deutsch)	negativ

Tabelle 5.9: IV-Verfahren – Modell 4.

IV-Modell 5 – Berufliche Vergangenheit. Ist ein Individuum im erlernten Beruf tätig, kann von einer längeren Ausbildungsdauer ausgegangen werden, da die Substitution bei Geringqualifizierten zumeist leichter und auch notwendiger ist. Der Einfluss dieser Variablen wird im nächsten Schritt mitberücksichtigt, das Modell lautet nun:

$$S_i = \gamma_0 + \gamma_1 D_{MB_i} + \gamma_2 D_{HB_i} + \gamma_3 N Buecher_i + \gamma_4 N Schw_i + \gamma_5 N Brued_i \\ + \gamma_6 D_{Nicht-Deusch_i} + \gamma_7 D_{Tätiger_i} + \sum_{j>7}^{15} \gamma_j X_{ij} + \eta_i,$$

wobei die Variable in Tabelle 5.10 näher dargestellt wird.

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
$D_{Tätiger_i}$	Tätigkeit im erlernten Beruf	dichotom (1: erlernt, 0: sonst)	positiv

Tabelle 5.10: IV-Verfahren – Modell 5.

Nach Prüfung der Voraussetzungen für das IV-Verfahren, dargestellt in Tabelle 5.13, ergibt sich, dass das Modell unter Berücksichtigung der beruflichen Vergangenheit zur Ablehnung des F-Tests auf die Residuen führt und damit die entscheidende Voraussetzung des Modells nicht erfüllt.¹³⁸ Damit erweist sich die Hinzunahme dieser Variablen als nicht geeignet. Im nächsten Modell wird sie daher nicht weiter berücksichtigt.

IV-Modell 6 – Politische Veränderungen. In Deutschland muss auf die politischen Veränderungen eingegangen werden. So kann der Bildungsabschluss aus der ehemaligen DDR umgerechnet und annähernd vergleichbar gemacht werden. Unberücksichtigt bleiben die verschiedenen Anreizsysteme zwischen Ost- und Westdeutschland vor der Wende. Daher wird eine Dummy-Variable für den ostdeutschen Schulabschluss in das Modell einbezogen. Die Spezifikation lautet nun:

$$S_i = \gamma_0 + \gamma_1 D_{MB_i} + \gamma_2 D_{HB_i} + \gamma_3 N Buecher_i + \gamma_4 N Schw_i + \gamma_5 N Brued_i \\ + \gamma_6 D_{Nicht-Deusch_i} + \gamma_7 D_{DDR} + \sum_{j>7}^{15} \gamma_j X_{ij} + \eta_i,$$

wobei die Variablenbeschreibung in Tabelle 5.11 gezeigt wird.

¹³⁸Näheres wird in Abschnitt 5.2.1 diskutiert.

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
D_{DDR}	Abschluss aus der DDR	dichotom (1: Abschluss aus der DDR, 0: sonst)	?

Tabelle 5.11: IV-Verfahren – Modell 6.

Aus der Geschichte kann vermutet werden, dass es im ehemaligen Arbeiterstaat der DDR geringere Anreize gab, sich höher zu qualifizieren. Demzufolge würde ein negativer Zusammenhang zwischen dieser Variablen und der Bildungsdauer erwartet werden. Betrachtet wird hierbei jedoch ein Datensatz, der ausschließlich Arbeitnehmer berücksichtigt. In Abschnitt 4.2.2 ist bereits gezeigt, dass im Jahr 2008 der Anteil der hochqualifizierten Arbeitnehmer aus der ehemaligen DDR deutlich höher ist als dies beim Anteil der westdeutschen Bevölkerung der Fall ist. Die Neuorientierung nach der Wende war vermutlich besonders für Geringqualifizierte in Ostdeutschland erschwert, sodass deren Anteil unter den Arbeitnehmern deutlich geringer ausfällt. Dies könnte daher zu einem positiven Koeffizienten führen. Demnach kann der vermutete Zusammenhang mit dem Schulabschluss nicht eindeutig festgestellt werden und ergibt sich erst aus der Regression. Auch diese Variable trägt zu einer Korrelation¹³⁹ mit dem Fehlerterm der Lohngleichung bei und ist demnach als Instrument nicht geeignet. Die detaillierte Darstellung und Prüfung der Modelle folgt im nächsten Teilabschnitt.

Voraussetzungen und Prüfungen der IV-Regression. Um die IV-Methode anwenden zu können, müssen folgende Voraussetzungen¹⁴⁰ erfüllt sein:

IV – V1 *Eliminierung des verzerrenden Einflusses der endogenen Variablen. Die Instrumente der Reduced-Form-Regression dürfen nicht mit dem Fehlerterm der Lohngleichung korreliert sein, d. h. $E(z'_i, \varepsilon_i) = 0$, wobei z'_i den Vektor der Instrumente beschreibt und ε_i für das Residuum der Lohngleichung steht.*

Bound et al. (1995) schlagen zur Prüfung des Erwartungswertes zwischen Instrumenten und Residuum der Lohngleichung einen F-Test aus einer weiteren Regression vor. Dabei werden die Instrumente auf die Residuen der Lohngleichung ge-

¹³⁹Siehe Tabelle 5.13.

¹⁴⁰Siehe hierzu Wooldridge (2002) und Murray (2006).

schätzt. Dies hat das Ziel zu prüfen, ob die zuvor kritisierte Endogenität der Bildungsvariable durch die Instrumente eliminiert werden konnte. Diese drückte sich aus über eine Korrelation zwischen der endogenen Variable mit dem Fehlerterm der Lohngleichung. Über den F-Test wird nun die Nullhypothese, ob die Variablen der Bildungsgleichung keinen signifikanten Erklärungsbeitrag für die Residuen der Lohngleichung liefern, geprüft. Ist dies erfüllt, so kann davon ausgegangen werden, dass der verursachende Effekt in den Instrumenten nicht mehr enthalten ist und demnach zu einer neu geschätzten Bildungsdauer führt, die exogen ist. Die detaillierte Beschreibung befindet sich in Anhang A.3.3 in Tabelle A.12. Die P-Werte sind pro Modellspezifikation in Tabelle 5.13 in der Zeile „F-Test auf Residuen“ angegeben. Abgesehen von den Modellen 5 und 6 kann die Nullhypothese auf dem 5 %-Signifikanzniveau nicht abgelehnt werden. Die Modelle 1–4 erfüllen damit diese Voraussetzung.

IV – V2 *Relevanz der Instrumentvariablen. Die Instrumentvariablen liefern einen Erklärungsbeitrag für die endogene Variable, d. h. $E(x_i, z_i') \neq 0$, wobei x_i die endogene Variable beschreibt.*

Zur Prüfung der Relevanz der Instrumente wird die F-Statistik der Reduced-Form-Regression genutzt. Kann die Nullhypothese der Nichtrelevanz abgelehnt werden, ist davon auszugehen, dass die Instrumente grundsätzlich einen Erklärungsbeitrag für die Bildungsdauer liefern. Die P-Werte für diesen Test sind in der Ergebnistabelle 5.13 pro Modell angegeben. Alle Modelle lassen die Ablehnung der Nullhypothese zu.

Mit welchem Anteil das Modell die Streuung der Bildungsdauer erklären kann, wird durch das Bestimmtheitsmaß aus der Reduced-Form-Regression gezeigt. Dieses dient als Kriterium zur Ermittlung des besten Modells. Diese Herangehensweise über die beiden F-Tests und den Vergleich des R^2 können als guter Indikator zur Beurteilung der IV-Schätzung betrachtet werden. So schreibt Bound et al. (1995) zur Beurteilung der beiden genannten Kriterien: „[...] the partial R^2 and F statistic on the excluded instrument in the first-stage regression are useful as rough guides to the quality of IV estimates.“

IV – V3 *Die Instrumente sind nicht perfekt multikollinear.*

Die Multikollinearitätsprüfung wurde bereits für die OLS-Regression durchgeführt. Demzufolge ist die nähere Verfahrenbeschreibung bereits in Abschnitt 5.1 gezeigt.

Die Darstellung der VIF pro Modellspezifikation für die Reduced-Form-Regression findet sich im Anhang A.3.3 in Tabelle A.14. Es zeigt sich eine geringe Multikollinearität zwischen den Instrumentvariablen, die deutlich unterhalb des kritischen Wertes¹⁴¹ liegt. Damit kann davon ausgegangen werden, dass die entstehenden Verzerrungen von marginaler Bedeutung sind und die Modelle aufgrund dieser Annahme alle akzeptiert werden können.

IV – V4 *Der Rang der Reduced-Form-Regression ist größer oder gleich der Schätzgleichung für den Lohn.*

Dies bedeutet, dass es zumindest so viele Instrumente wie endogene Variablen¹⁴² gibt. Allgemein gilt, dass das Modell exakt-, über- bzw. unteridentifiziert sein kann. Liegt exakte Identifizierung vor, so ist die Anzahl der Instrumente genau gleich der Anzahl der endogenen Regressoren. Bei Überidentifizierung sind mehr Instrumente als endogene Regressoren vorhanden und umgekehrt bei unteridentifizierten Modellen. Problematisch wird die Analyse für letztgenannte, da diese nicht schätzbar sind.¹⁴³ Es ist festzustellen, dass sich lediglich eine Zusatzregression ergibt, die mit 2 bis 7 Instrumentvariablen zu schätzen ist. Mit diesen Modellspezifikationen liegen überidentifizierte Modelle vor.

IV – V5 *Die Fehlerterme der Reduced-Form-Regression und der Ursprungsregression dürfen nicht korreliert sein. Damit gilt, dass $cov(\varepsilon_i, \eta_i) = 0$. Liegt doch ein Zusammenhang vor, spricht man von kontemporärer Korrelation.*

In Tabelle 5.12 werden die jeweiligen Korrelationskoeffizienten zwischen den Residuen der Lohn- und Bildungsgleichung dargestellt. Für alle Modelle muss diese Voraussetzung abgelehnt werden. Eine Korrelation ist für jede Modellspezifikation auf dem 5 %-Signifikanzniveau vorhanden. Da diese Voraussetzung verletzt ist, bietet sich anstelle der sequentiellen Schätzung eine simultane an. Sie wird über das Three-Stage Least Squares Verfahren (3SLS-Verfahren) vorgenommen. Zunächst werden die Modellergebnisse des IV-Verfahrens gezeigt, um das beste Modell zur Schätzung der Bildungsdauer zu identifizieren. Im nächsten Schritt wird dieses Modell im 3SLS-Verfahren angewendet.

¹⁴¹ Siehe hierzu Abschnitt: 5.1, Voraussetzung 6.

¹⁴² Vgl. Wooldridge (2002), S. 93 f.

¹⁴³ Siehe hierzu Abschnitt 5.2.1.

	Korrelationskoeffizient zwischen den Residuen
IV – Modell 1	-0,069***
IV – Modell 2	-0,128***
IV – Modell 3	-0,109***
IV – Modell 4	-0,091***
IV – Modell 5	-0,200***
IV – Modell 6	-0,033**

Tabelle 5.12: Korrelationen zwischen dem Residuum der Lohn- und Bildungsgleichung pro Modellspezifikation; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %.

Ergebnisdarstellung des IV-Verfahrens. In diesem Teilabschnitt werden die Ergebnisse aus der Regression auf die Bildungsdauer dargestellt. Dabei wird der besondere Fokus auf die Erfüllung der Voraussetzungen des IV-Verfahrens gelegt. Erfüllt die Hinzunahme der Variablen die Voraussetzungen zur Lösung des Endogenitätsproblems, so wird sie im weiteren Modell berücksichtigt, andernfalls nicht.

Ergebnisse der 1. Stufe (Reduced-Form-Regression). In den Tabellen 5.13 und 5.14 sind die Ergebnisse der Reduced-Form-Regression gezeigt. Dabei werden die einzelnen Schätzergebnisse pro Modellspezifikation dargestellt. Die beschriebenen Instrumente haben einen signifikanten Einfluss auf die Bildungsdauer, unabhängig von der gewählten Modellspezifikation. Der besondere Einfluss der väterlichen Bildung kann deutlich bestätigt werden, ebenso das anhand der vorhandenen Bücher im Haushalt unterstellte intellektuelle Niveau. Ein Substitutionsverhältnis zwischen der Qualität der Bildung der Kinder und ihrer Anzahl ist auch empirisch zu belegen. So wird für die Zahl der Schwestern und Brüder je ein negativer Koeffizient geschätzt. Der Einfluss der deutschen Staatsbürgerschaft auf die Bildungsdauer ist ebenso sehr deutlich zu zeigen. Eine Person mit ausländischer Staatsbürgerschaft hat im Schnitt eine deutlich geringere Qualifikation. Im Gegensatz dazu hat eine Person, die in ihrem Ausbildungsberuf tatsächlich tätig ist, tendenziell eine höhere Bildungsdauer als ein Quereinsteiger. Ebenso haben Personen mit ostdeutschem Abschluss eher eine höhere Bildungsdauer. Dieses Ergebnis, entgegen der Intuition, ist aufgrund der alleinigen Berücksichtigung von Individuen mit positivem Einkommen zu erklären.

Zur Auswahl des geeigneten Modells muss die Prüfung der Annahmen herangezogen werden. Die Voraussetzung (IV – V1) ist für die Modelle 1 bis 4 annehmbar. Bei den gegebenen P-Werten besteht keine empirische Evidenz zur Ablehnung der Nullhypothese der

5 Schätzmethoden und Ermittlung der nationalen Bildungsrendite

Variable	IV – Modell 1	IV – Modell 2	IV – Modell 3	IV – Modell 4	IV – Modell 5	IV – Modell 6
D_{MB_i}	1,793*** [0,097]	1,536*** [0,092]	1,482*** [0,091]	1,459*** [0,090]	1,439*** [0,088]	1,442*** [0,090]
D_{HB_i}	4,696*** [0,126]	3,961*** [0,130]	3,917*** [0,129]	3,894*** [0,129]	3,797*** [0,129]	3,853*** [0,129]
N_{Buec}		1,684*** [0,077]	1,644*** [0,076]	1,581*** [0,076]	1,476*** [0,075]	1,598*** [0,076]
N_{Schw}			-0,155*** [0,032]	-0,129*** [0,031]	-0,105*** [0,030]	-0,125*** [0,031]
N_{Brud}			-0,205*** [0,029]	-0,183*** [0,028]	-0,174*** [0,027]	-0,175*** [0,028]
$D_{nichtdeu}$				-1,250*** [0,139]	-1,050*** [0,134]	-1,166*** [0,140]
$D_{taetierl}$					0,930*** [0,064]	
D_{DDR}						0,387*** [0,074]
Erf	-0,062*** [0,021]	-0,074*** [0,020]	-0,062*** [0,020]	-0,071*** [0,019]	-0,058*** [0,019]	-0,082*** [0,019]
Erf^2	-0,000 [0,000]	0,000 [0,000]	-0,000 [0,000]	0,000 [0,000]	-0,000 [0,000]	0,000 [0,000]
$weibl$	0,295** [0,121]	0,246** [0,116]	0,248** [0,115]	0,268** [0,115]	0,222** [0,113]	0,256** [0,114]
$verhei$	0,400*** [0,106]	0,341*** [0,101]	0,368*** [0,100]	0,444*** [0,100]	0,414*** [0,098]	0,452*** [0,100]
IAT_{wver}	-0,261* [0,145]	-0,257* [0,138]	-0,264* [0,137]	-0,318** [0,136]	-0,274** [0,133]	-0,325** [0,136]
D_{groB}	0,174*** [0,068]	0,125** [0,064]	0,098 [0,063]	0,105* [0,063]	0,128** [0,061]	0,128** [0,063]
D_{laendl}	-0,197*** [0,069]	-0,155** [0,065]	-0,138** [0,065]	-0,181*** [0,064]	-0,161** [0,063]	-0,216*** [0,065]
D_{Pendle}	0,088 [0,069]	0,110* [0,065]	0,095 [0,064]	0,074 [0,064]	0,051 [0,062]	0,091 [0,064]
$Konst$	12,933*** [0,335]	12,850*** [0,320]	12,999*** [0,318]	14,382*** [0,349]	13,457*** [0,341]	14,346*** [0,349]

Tabelle 5.13: Ergebnistabelle der Reduced-Form-Regression Teil 1; in eckigen Klammern sind die Standardfehler nach White mit den Signifikanzniveaus: *** 1 %, ** 5 % und * 10 % angegeben; Quelle: DIW, eigene Berechnungen.

Variable	IV – Modell 1	IV – Modell 2	IV – Modell 3	IV – Modell 4	IV – Modell 5	IV – Modell 6
korr. R ²	0,30	0,38	0,39	0,40	0,43	0,40
F-Test auf Residuen	1,000	0,545	0,523	0,571	0,000	0,000
F-Test	0,000	0,000	0,000	0,000	0,000	0,000
JB-Test	0,000	0,000	0,000	0,000	0,000	0,000
n	4.693	4.693	4.693	4.693	4.693	4.693

Tabelle 5.14: Ergebnistabelle der Reduced-Form-Regression Teil 2. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

gegebenen Korrelation der Instrumente mit dem Fehlerterm der Lohngleichung. Durch die Hinzunahme der Variablen Tätigkeit im erlernten Beruf und DDR-Abschluss ergibt sich in den Modellen 5 und 6 eine Korrelation zum Fehlerterm der Lohngleichung. Dies zeigt sich durch die signifikante Ablehnung der Nullhypothese des F-Tests auf die Residuen. Alle Modelle zeigen den geforderten Erklärungszusammenhang in Voraussetzung 2 (IV – V2). Die Instrumentvariablen liefern demnach einen Erklärungsbeitrag für die Bildungsdauer. Die Modelle sind geeignet um diese Variable zu erklären. Die Multikollinearitätsprüfung ist im Anhang in Tabelle A.14 dargestellt. Gemäß der sich ergebenden VIF kann diese Problematik als gering eingeschätzt werden.

Da die Voraussetzung 5 (IV – V5) für kein Modell bestätigt werden kann, ist zunächst das beste Modell aus den Spezifikationen 1 bis 4 auszuwählen. Bei diesen vier Modellen sind zumindest alle anderen Voraussetzungen erfüllt. Hierfür wird die Güte des Modells (IV – V2) der Reduced-Form-Regression den Ausschlag geben. Als das beste Modell kann dabei die Spezifikation 4 bestimmt werden. Dieses mit einem R² von 40 % geschätzte Modell wird für die weiteren Verfahren genutzt. Zudem werden die sich ergebenden P-Werte des Jarque-Bera-Test¹⁴⁴ gezeigt. Wie bereits bei den OLS-Regressionen muss die Nullhypothese der Homoskeastizität für alle Modelle abgelehnt werden. Daher werden die korrigierten Standardfehler von White für die Signifikanztests verwendet. Die geschätzten Werte für die Bildungsdauer aus der Reduced-Form-Regression werden nun in die Lohngleichung als neuer exogener Regressor implementiert und geschätzt.

Ergebnisdarstellung der 2. Stufe (Schätzung der Lohngleichung). Die Schätzung des 2. Schritts des IV-Verfahrens wird an dieser Stelle diskutiert und in Tabelle 5.15

¹⁴⁴Die nähere Beschreibung findet sich in Abschnitt 5.1.

5 Schätzmethoden und Ermittlung der nationalen Bildungsrendite

dargestellt. Vergleicht man die verschiedenen IV-Modelle und ihre Auswirkungen auf die Bildungsrendite, so zeigen die Modelle 2, 3 und 4 quasi identische Koeffizienten. Wie bereits durch die F-Statistiken gezeigt, ergibt sich kaum eine Änderung durch die Hinzunahme von einzelnen Instrumentvariablen bei den Modellspezifikationen 1 bis 4.

Variable	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5	Modell 6
Bildungsdauer	0,092*** [0,005]	0,100*** [0,004]	0,099*** [0,004]	0,099*** [0,004]	0,109*** [0,004]	0,090*** [0,004]
Erfahrung	0,030*** [0,004]	0,032*** [0,004]	0,032*** [0,004]	0,032*** [0,004]	0,033*** [0,004]	0,030*** [0,004]
Erfahrung ²	-0,000*** [0,000]	-0,000*** [0,000]	-0,000*** [0,000]	-0,000*** [0,000]	-0,000*** [0,000]	-0,000*** [0,000]
weiblich	-0,127*** [0,021]	-0,130*** [0,021]	-0,130*** [0,021]	-0,130*** [0,021]	-0,133*** [0,021]	-0,127*** [0,021]
verheiratet	0,111*** [0,018]	0,108*** [0,018]	0,108*** [0,018]	0,108*** [0,018]	0,104*** [0,018]	0,112*** [0,018]
IAT verh. Frau	-0,142*** [0,025]	-0,140*** [0,025]	-0,140*** [0,025]	-0,140*** [0,025]	-0,137*** [0,025]	-0,142*** [0,025]
großes Untern. ländliches Gebiet	0,218*** [0,012]	0,216*** [0,012]	0,216*** [0,012]	0,216*** [0,012]	0,213*** [0,012]	0,219*** [0,012]
Pendler	-0,050*** [0,012]	-0,048*** [0,012]	-0,048*** [0,012]	-0,048*** [0,012]	-0,045*** [0,012]	-0,051*** [0,012]
Konstante	0,063*** [0,012]	0,063*** [0,012]	0,063*** [0,012]	0,063*** [0,012]	0,063*** [0,012]	0,063*** [0,012]
	1,113*** [0,090]	0,996*** [0,079]	1,002*** [0,078]	1,009*** [0,078]	0,857*** [0,078]	1,140*** [0,077]
R^2	0,35	0,34	0,34	0,34	0,33	0,35
Wald- Test	0,000	0,000	0,000	0,000	0,000	0,000
BP-Test	0,000	0,000	0,000	0,000	0,000	0,000
n	4.693	4.693	4.693	4.693	4.693	4.693

Tabelle 5.15: Ergebnisse für die Schätzung der Lohngleichung mit der IV-Methode; in eckigen Klammern sind die Standardfehler nach White mit den Signifikanzniveaus: *** 1 %, ** 5 % und * 10 % angegeben. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

Da die Validität der Instrumente für die Modelle 5 und 6 aufgrund einer Korrelation mit dem Fehlerterm der Lohngleichung abgelehnt wurde, sind diese Modellergebnisse an dieser Stelle lediglich der Vollständigkeit halber angegeben. Mit diesem IV-Ansatz kann eine

durchschnittlich jährliche Verzinsung eines zusätzlichen Bildungsjahres von ca. 10,4 %¹⁴⁵ ermittelt werden. Der niedrigste Koeffizient aus den möglichen IV-Spezifikationen ergibt sich aus Modell 1. Hierbei wird eine durchschnittliche jährliche Verzinsung von ca. 9,6 % pro zusätzlichem Bildungsjahr berechnet. Der Vergleich mit den OLS-Ergebnissen zeigt in allen 4 Modellspezifikationen eine Steigerung der Renditen durch die Verwendung des zweistufigen IV-Verfahrens.

Die weiteren Koeffizienten bestätigen den in den Modellbeschreibungen in Abschnitt 5.1.1 vermuteten Zusammenhang zum Bruttostundenlohn. Sie sind mit einer Irrtumswahrscheinlichkeit von 5 % signifikant von null verschieden. Des Weiteren sind im Vergleich die Koeffizienten der Kontrollvariablen der Modelle 1 bis 4 unabhängig von der Modellspezifikation der Reduced-Form-Regression stabil. Dementsprechend haben die Berufserfahrung (in linearer Form), der Familienstand, das Arbeiten in großen Unternehmen sowie das Pendlerverhalten einen positiven Einfluss auf den logarithmierten Bruttostundenlohn. Im Gegensatz dazu wirken die quadrierte Berufserfahrung, das weibliche Geschlecht, der Ehestand der Frau sowie der Wohnort in einem ländlichen Gebiet negativ auf das Einkommen.

Die Güte des Regressionsmodells unterscheidet sich zwischen den 4 möglichen IV-Modellspezifikationen kaum. Das R^2 zeigt, dass stets zwischen 34 % und 35 % der Gesamtstreuung des logarithmierten Bruttostundenlohn erklärt werden können. Für die Schätzung wird das allgemeinere GLS-Verfahren¹⁴⁶ angewendet. Hierfür wird zur Beurteilung der Gesamtsignifikanz anstelle des F-Tests der Wald-Test verwendet. Dieser gibt ebenso eine Hilfestellung zur Beurteilung des Erklärungsgehalts des Gesamtmodells für den Regressand, basiert jedoch auf einer χ^2 -Verteilung. Alle Modelle zeigen hierbei, dass die Spezifikation einen Erklärungsbeitrag für den logarithmierten Bruttostundenlohn liefert. Der Test auf Homoskedastizität muss, wie bei den OLS-Regressionen, für alle Spezifikationen abgelehnt werden, sodass die robusten Standardfehler von White zur Beurteilung der Signifikanz der Koeffizienten Eingang finden.

Vergleich der Ergebnisse. Die Bildungsrendite des besten Modells (Spezifikation 4) aus dem OLS-Verfahren ergibt eine mittlere jährliche Verzinsung von 8,5 % pro zusätzlichem Bildungsjahr. Im Vergleich kann mit dem besten Modell (IV – Modell 4) aus dem IV-Verfahren eine Rendite in Höhe von 10,4 % ermittelt werden. Das Ergebnis aus der

¹⁴⁵Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs stets wie folgt berechnet:
$$\text{prozentuale Bildungsrendite} = (e^{\beta_1} - 1) \cdot 100.$$

¹⁴⁶Generalized Least Squares. Näheres zu diesem Test in Heij et al. (2004), S. 242 ff.

OLS-Regression muss als verzerrt angenommen werden, da nachweislich die Variable der Bildungsdauer das Exogenitätskriterium nicht erfüllt. Für das IV-Verfahren wurden valide Instrumente gefunden. Die statistischen Tests auf die Validität und den Erklärungsgehalt dieser Instrumentierung lassen auf eine sinnvolle Schätzgleichung schließen. Die Voraussetzung der unkorrelierten Störterme der Lohn- und Bildungsgleichung (IV – V5) ist nicht erfüllt. Demnach muss davon ausgegangen werden, dass es systematische unberücksichtigte Größen gibt, die auf beide Fehlerterme gleichermaßen wirken. Dem Problem dieser sogenannten kontemporären Korrelation wird in Abschnitt 5.2.2 begegnet.

Kritik. Die Annahmenprüfung ergibt, dass das Endogenitätsproblem gelöst werden kann. Die Annahme einer fehlenden Korrelation der Fehlerterme zwischen Lohn- und Bildungsgleichung ist hingegen nicht erfüllt. Dies kann durch eine simultane Schätzung der Modelle mit dem Three-Stage Least Squares Verfahren gelöst werden. Im nachfolgenden Abschnitt wird dies diskutiert.

Betrachtet man die Veränderung der Bildungsrenditen aus der IV- und OLS- Schätzung, so ergibt sich, dass der geschätzte Koeffizient größer ist als zuvor. Ziel der IV- Schätzung ist es, die Verzerrung durch die mangelnde Quantifizierbarkeit der Variablen Motivation und Fähigkeit zu eliminieren. Intuitiv klar ist, dass eine höhere Motivation positiv auf das Einkommen wirkt. Daraus würde man einen überschätzten Koeffizienten für die OLS-Modelle vermuten. Demnach müssten die IV- Schätzer der Bildungsrendite einen geringeren Wert aufweisen. Bei den geschätzten Modellen dieser Arbeit sowie auch in fast allen anderen Untersuchungen¹⁴⁷ ist das Gegenteil der Fall. Card (2001) bemerkt hierzu, dass „*the findings from the recent IV literature are remarkably consistent with Griliches' (1977) [...], and point to a causal effect of education that is as big or bigger than the OLS estimated return [...]*“¹⁴⁸. Durch das Herausrechnen der Motivation kann der Einfluss der Bildungsjahre auf das Einkommen gesteigert werden. Diverse Untersuchungen diskutieren dieses kontraintuitive Ergebnis.

- Griliches (1977) und Angrist und Krueger (2001) zeigen, dass der Unterschied zwischen OLS- und IV-Schätzern für die Rendite relativ gering ist. Diese geringe Unterschätzung, die gefunden werden müsste, wird kompensiert durch entstehende Messfehler. Dabei kann zum einen die fehlende Variable in der Lohngleichung beim OLS-Verfahren zu Verzerrungen führen. Auf der anderen Seite ist auch die Auswahl

¹⁴⁷Einige sind in Abschnitt 3 beschrieben.

¹⁴⁸Card (2001), S. 1157 f.

der Instrumente begrenzt und es kann zu Messfehlern auf Seiten der zu schätzenden Bildungsdauer kommen. Diese Fehler führen zur Erhöhung des Schätzers mit der IV-Methode im Vergleich zur Anwendung der OLS-Methode.

- Ashenfelter et al. (1999) begründen diese Unterschiede durch die bewusste Auswahl der Instrumente. Dabei werden Instrumente als geeignet eingestuft, wenn sie u. a. eine gute t-Statistik für die Bildungsrendite liefern, bzw. werden jene ohne signifikante Bildungsrendite von vornherein aussortiert, um ein publizierbares Ergebnis zu präsentieren.
- Fraglich ist jedoch, ob die Ergebnisse aus der OLS-Regression überhaupt als Vergleich herangezogen werden können. Aufgrund des bestätigten Endogenitätsproblems sind die Schätzer verzerrt. Damit scheint das generelle Heranziehen der OLS-Koeffizienten für die Bildungsrendite als fragwürdig. Andererseits sind im Literaturvergleich diese Schätzer zumeist sehr homogen und schwanken unabhängig von Datensatz und Modellspezifikation zwischen 7 % und 9 %.

Für die Interpretation der Bildungsrenditen muss beachtet werden, dass es sich bei Regressionsverfahren lediglich um geschätzte Koeffizienten handelt, die als Durchschnittswerte zu interpretieren sind und von denen eine absolute Genauigkeit leider nie erwartet werden kann. Die Ergebnisse zeigen ungefähre Werte an und müssen demnach als solche interpretiert werden.

5.2.2 Three-Stage Least Squares

Werden Mehrgleichungsmodelle sequentiell geschätzt, so kann das Problem der kontemporären Korrelation auftreten. Dies bedeutet, dass die Störterme der verschiedenen Regressionen miteinander korreliert sind. Für die Koeffizienten ergibt sich damit, dass die Schätzer zwar unverzerrt und konsistent, nicht jedoch effizient geschätzt werden. Um dem zu begegnen wird eine Methode nach Zellner und Theil (1962) gewählt, in der eine Kombination zwischen dem IV-Verfahren und einem Seemingly-Unrelated-Regressionsmodell¹⁴⁹ angewendet wird. Dafür gilt es, zunächst die IV-Methode separat zu schätzen. Die ersten beiden Schritte sind demnach identisch zum IV-Verfahren. Nun wird die Kovarianzmatrix¹⁵⁰ der Fehlerterme ermittelt. Im dritten Schritt wird das Gleichungssystem per

¹⁴⁹Kurz SUR-Modell. Siehe auch Heij et al. (2004), S. 685.

¹⁵⁰Die Kovarianzmatrix wird berechnet durch $s_{ij} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \eta_i$. Vgl. Heij et al. (2004), S. 686 und S. 706.

Feasible-Generalized Least Squares¹⁵¹ unter Hinzunahme der gemeinsamen Kovarianzmatrix simultan geschätzt.

Diesen Modellen sind jedoch aufgrund der Simultanität Grenzen gesetzt. So ist zu befürchten, dass bei geringen Stichprobengrößen die Kovarianzmatrix nicht richtig geschätzt werden kann. Ist eine der geschätzten Gleichungen fehlspezifiziert, so hat dies negative Auswirkungen auf die Konsistenz aller Schätzer. In beiden Fällen ist das 3SLS-Verfahren nachrangig zum IV-Verfahren zu bewerten. Eine Anwendung des 3SLS-Verfahrens ergibt damit nur dann einen Nutzen, wenn das Vertrauen in alle zu schätzenden Modellspezifikationen gegeben ist und der Stichprobenumfang ausreichend groß ist. In diesem Fall kann der Vorteil der gemeinsamen Kovarianzmatrix und der damit einhergehenden Bereinigung der kontemporären Korrelation genutzt werden. Getestet wird der Erklärungsgehalt des Gesamtmodells über den Wald-Test.¹⁵²

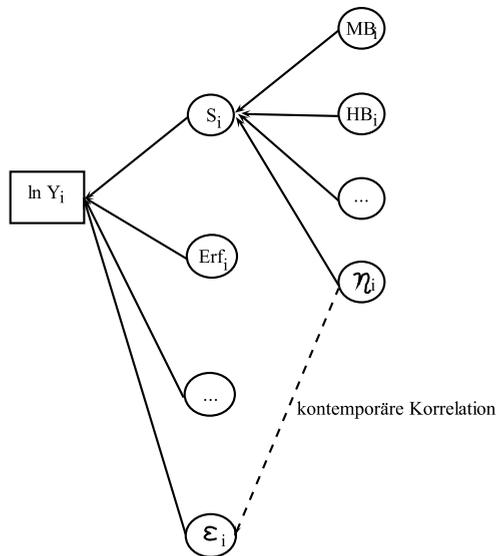


Abbildung 5.5: Kontemporäre Korrelation bei Mehrgleichungsmodellen; eigene Darstellung.

Bezug zur Bildungsrendite. Das IV-Verfahren wird zur Behebung der Endogenität der Variable der Bildungsdauer angewendet. Problematisch dabei ist, dass die Fehlerterme aus der sequentiellen Schätzung korreliert sind und demnach die Koeffizienten nicht effizient

¹⁵¹Kurz FGLS.

¹⁵²Siehe hierzu Dhrymes (1969).

geschätzt werden können. Abbildung 5.5 verdeutlicht diese Problematik näher. Für das Referenzmodell (Spezifikation 4) des IV-Verfahrens ist der Korrelationskoeffizient zwischen beiden Fehlertermen mit $\text{corr}(\varepsilon_i, \eta_i) = -0,091$ signifikant¹⁵³ von null verschieden. Alle weiteren Voraussetzungen des IV-Verfahrens sind bei diesem Modell erfüllt. Um die 3SLS-Methode anzuwenden, wird eine genügend große Stichprobe sowie das Vertrauen auf die richtige Modellspezifikation vorausgesetzt. Ersteres ist ohne Zweifel mit einem Stichprobenumfang von 4.693 Beobachtungen zu belegen. Auch die genaue Spezifikation kann durch die Prüfung der OLS- sowie IV-Voraussetzungen als vertrauenswürdig angesehen werden. Die Anwendung des 3SLS-Verfahrens ist somit ein geeigneter Weg zur Bereinigung der kontemporären Korrelation.

Ergebnisdarstellung. Durch die simultane Schätzung beider Regressionsmodelle unter Verwendung der gemeinsamen Kovarianzmatrix der Fehlerterme wird, anders als im Falle der IV-Schätzung, nur jeweils ein Koeffizient pro Kontrollvariable geschätzt. Bei Betrachtung der Regression des Lohngleichungsmodells ergibt sich eine Bildungsrendite¹⁵⁴ in Höhe von 10,2 %.

Wie auch in den anderen Verfahren können für alle Regressionskoeffizienten signifikante Ergebnisse geschätzt werden. Die Wirkung der Berufserfahrung ist positiv und im Vergleich zu den anderen Verfahren auch in der Höhe ähnlich. Negativ auf das Einkommen wirken sich weiterhin das weibliche Geschlecht, die Kombination für den Ehestand der Frau sowie ein Wohnort in ländlichem Gebiet aus. Positiven Einfluss haben der ehe- bzw. eheähnliche Familienstand, das Arbeiten in großen Unternehmen sowie Pendlerverhalten. Auch die Koeffizienten der Reduced-Form-Regression entsprechen den Erwartungen und sind signifikant von null verschieden. Positiv wirken demnach eine mittlere und hohe väterliche Bildung auf die eigene Bildung. Ebenso zeigt der Koeffizient für die Anzahl der Bücher im Haushalt ein positives Vorzeichen. Im Gegensatz dazu wirken die Anzahl der Geschwister sowie eine nicht deutsche Staatsangehörigkeit tendenziell negativ auf die Bildungsdauer.

Bei der Betrachtung der Inferenzstatistik ergibt sich für beide Regressionen eine ähnliche Güte der Modelle. So werden mit den gewählten Spezifikationen ca. 34 % von der Gesamtstreuung des logarithmierten Bruttostundenlohns und 36 % von der Bildungsdauer

¹⁵³Siehe Tabelle 5.12.

¹⁵⁴Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs stets wie folgt berechnet:
 $\text{prozentuale Bildungsrendite} = (e^{\beta_1} - 1) \cdot 100$.

erklärt. Der Wald-Test bestätigt auf dem 5 %-Signifikanz-niveau den grundsätzlichen Erklärungsbeitrag beider Modelle für den Regressand.

Instrumente	Koeffizient	allgemeine Regression	Koeffizient
mittlere Bildung des Vaters	1,729*** (0,073)	Bildungsjahre	0,097*** (0,004)
hohe Bildung des Vaters	4,346*** (0,140)	Erfahrung	0,030*** (0,004)
Anzahl der Bücher	1,548*** (0,071)	Erfahrung ²	-0,000*** (0,000)
Anzahl der Schwestern	-0,176*** (0,031)	weiblich	-0,123*** (0,021)
Anzahl der Brüder	-0,211*** (0,031)	verheiratet	0,119*** (0,018)
Staatsangeh. nicht deutsch	-1,106*** (0,143)	IAT verh. Frau	-0,147*** (0,025)
Konstante	13,036*** (0,159)	großes Unternehmen	0,218*** (0,012)
		ländliches Gebiet	-0,052*** (0,012)
		Pendler	0,065*** (0,012)
		Konstante	1,055*** (0,080)
korr. R ²	0,36	korr. R ²	0,34
Wald-Test	0,000	Wald-Test	0,000
n	4.693	n	4.693

Tabelle 5.16: Ergebnistabelle der 3SLS-Regression; Signifikanzniveau: *** 1 %, ** 5 % und * 10 % angegeben. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

5.2.3 Zusammenfassung

Der Vergleich zeigt, dass bei dem IV-Verfahren gegenüber dem OLS-Verfahren deutlich höhere Bildungsrenditen geschätzt werden. Das IV-Verfahren und das 3SLS-Verfahren unterscheiden sich lediglich hinsichtlich der Simultanität der Schätzung. Der sich ergebende Koeffizient für die Bildungsdauer ist nun mit 10,2 % dem aus dem IV-Verfahren von der Höhe ähnlich. Unter Berücksichtigung der Voraussetzungen für die jeweiligen Methoden kann das 3SLS-Verfahren als ein sinnvolles Verfahren zur Schätzung von Bildungsrenditen-

ten identifiziert werden. Der Unterschied zum IV-Verfahren ist nun, dass nicht nur von unverzerrten und konsistenten Schätzern ausgegangen werden kann, sondern zudem das Effizienzkriterium ebenso erfüllt wird, da um die kontemporäre Korrelation bereinigt werden kann.

Dieser Abschnitt zeigte eine ausführliche Diskussion verschiedener Verfahren und Modellspezifikationen mit einem prozessorientierten Aufbau des nachfolgend weiterhin verwendeten Modells und Verfahrens. Die ermittelte gesamtdeutsche Bildungsrendite kann mit der besten Modellspezifikation und dem geeignetsten Verfahren auf 10,2 % beziffert werden.

5.3 Bildungsrenditen nach verschiedenen Merkmalen

Im Weiteren werden in einer weiteren nationalen Betrachtung unterschiedliche Bildungsrenditen für ausgewählte Subgruppen berechnet. Dafür wird im Folgenden der Datensatz nach einzelnen Kriterien separat analysiert. Das in Abschnitt 5.1.1 bereits beschriebene Referenzmodell zur Schätzung der Lohnleichung lautet weiterhin:

$$\begin{aligned} \ln Y_{ik} &= \beta_{0k} + \beta_{1k} S_{ik} + \beta_{2k} Erf_{ik} + \beta_{3k} Erf_{ik}^2 + \beta_{4k} weibl_{ik} + \beta_{5k} verhei_{ik} \\ &+ \beta_{6k} IAT_{w-verh_{ik}} + \beta_{7k} D_{gro\beta_{ik}} + \beta_{8k} D_{ländl_{ik}} + \beta_{9k} D_{pendler_{ik}} + \epsilon_{ik}, \end{aligned}$$

ebenso wie die Variablenbezeichnung aus Abschnitt 5.1.1. Der Index i steht für das jeweilige Individuum. Neu hinzugefügt ist der Index k , der die k -te Gruppe des betrachteten Merkmals bestimmt. Im Speziellen werden separierte Bildungsrenditen für drei verschiedene Kohorten, für den öffentlichen Dienst und für die Privatwirtschaft sowie für voll- und teilzeitbeschäftigte Arbeitnehmer, berechnet. In Abschnitt 4.2.3 sind vergleichende deskriptive Statistiken zu den jeweiligen Untergruppen dargestellt. Diese lassen Unterschiede in der Bildungsrendite vermuten. Aus den vorangegangenen Abschnitten wird deutlich, dass sinnvolle Ergebnisse über das 3SLS-Verfahren ermittelt werden können. Das Referenzmodell für die Reduced-Form-Regression ist in Abschnitt 5.2.1 ausführlich diskutiert und wird identifiziert als:

$$S_{ik} = \gamma_{0k} + \gamma_{1k}D_{MB_{ik}} + \gamma_{2k}D_{HB_{ik}} + \gamma_{3k}NBuecher_{ik} + \gamma_{4k}NSchw_{ik} \\ + \gamma_{5k}NBrued_{ik} + \gamma_{6k}DNicht-Deutsch_{ik} + \eta_{ik}.$$

Für die einzelnen Untergruppen werden die Schätzungen der Lohngleichung und Bildungsdauer mit dem beschriebenen 3SLS-Verfahren vorgenommen.

5.3.1 Kohorteneffekte

Zu der jüngsten Kohorte gehören Personen zwischen dem 30. und 40. Lebensjahr. Es folgt das Alterscluster der 41–50-Jährigen sowie das der 51–60-Jährigen. Die deskriptive Analyse aus Abschnitt 4.2.3 zeigt bereits einen positiven Einkommenseffekt mit steigendem Alterscluster. Bei einer bivariaten Betrachtung kann der Einfluss einer höheren Berufserfahrung nicht herausgerechnet werden. Daher wird in einer multiplen Regression diese Variable als Kontrollvariable hinzugefügt.

Ergebnisdarstellung. Die sich ergebenden Koeffizienten für die Regressionen auf den logarithmierten Bruttostundenlohn sind in Tabelle 5.17 dargestellt. Die Ergebnisse aus der Regression auf die Bildungsdauer werden im Anhang in Tabelle A.15 präsentiert. Aufgrund der gruppenspezifischen Separierung des Datensatzes sind die einzelnen Stichprobenumfänge geringer, jedoch ergibt sich keine Gleichverteilung auf die Kohorten. Diese geringeren Beobachtungszahlen können als ein Grund für zum Teil insignifikante Koeffizienten angegeben werden. Die Bildungsrendite kann jedoch für alle Altersgruppen mit einer Sicherheit von 95 % als signifikant bestätigt werden. Je höher die Altersgruppe, umso höher ist die mittlere jährliche Verzinsung von Bildung. Somit kann für die 30–40-Jährigen eine Rendite von 10 %¹⁵⁵ und in der mittleren Kohorte von 11,2 % ermittelt werden. Für die älteste Kohorte wird eine mittlere Rendite von 13,1 % geschätzt. Mit dem Vergleich der Renditen ist noch nicht klar, ob die Unterschiede tatsächlich signifikant sind. Aufgrund der separierten Schätzung pro Subgruppe und demzufolge verschiedenen Varianz-Kovarianz-Matrizen kann ein statistischer Test nicht als Vergleichsinstrument herangezogen werden. Um jedoch trotzdem ein Indiz für die signifikante Unterschiedlichkeit zu erhalten, wer-

¹⁵⁵Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs wie folgt berechnet:
 $prozentuale\ Bildungsrendite = (e^{\beta_1} - 1) \cdot 100.$

5.3 Bildungsrenditen nach verschiedenen Merkmalen

den die 95 %-Konfidenzintervalle der geschätzten Bildungsrenditen verglichen. Dargestellt sind diese in Tabelle 5.18. Überschneiden sich diese Intervalle, so kann nicht davon ausgegangen werden, dass die Koeffizienten tatsächlich verschieden sind. Überschneiden sie sich umgekehrt dagegen nicht, ist dies ein Indiz für unterschiedliche Bildungsrenditen pro Altersgruppe. Aufgrund der gegebenen Intervalle kann nicht – zumindest bei einer Irrtumswahrscheinlichkeit von 5 % – von einer tatsächlich verschiedenen Bildungsrendite für die Altersgruppen ausgegangen werden.

Lohngleichung	30–40 Jahre	41–50 Jahre	51–60 Jahre
Bildungsdauer	0,095*** (0,011)	0,106*** (0,008)	0,123*** (0,000)
Erfahrung	0,046*** (0,017)	0,076*** (0,022)	0,053* (0,029)
Erfahrung ²	-0,001 (0,000)	-0,001*** (0,000)	-0,000 (0,000)
weiblich	-0,047 (0,031)	-0,195*** (0,032)	-0,148*** (0,050)
verheiratet	0,152*** (0,029)	0,099*** (0,028)	0,065 (0,041)
IAT verh. Frau	-0,219*** (0,041)	-0,137*** (0,038)	-0,043 (0,055)
großes Unternehmen	0,187*** (0,021)	0,233*** (0,017)	0,229*** (0,023)
ländliches Gebiet	-0,007 (0,021)	-0,065*** (0,018)	-0,077*** (0,023)
Pendler	0,055*** (0,021)	0,063*** (0,018)	0,070*** (0,023)
Konstante	0,791*** (0,251)	0,405 (0,344)	-0,041 (0,571)
korr. R ²	0,31	0,38	0,35
Wald-Test	0,000	0,000	0,000
n	1.410	1.871	1.412

Tabelle 5.17: Ergebnistabelle der 3SLS-Regression für die Kohorten; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

Für den Einfluss der Berufserfahrung wird aus der Theorie von Mincer (1974) ein positiv abnehmendes Grenzeinkommen erwartet. Dies ist mit Blick auf die älteste Gruppe und den sich ergebenden Koeffizienten erkennbar. Der Einfluss kann für diese Gruppe mit einer Sicherheit von 95 % als nicht signifikant angenommen werden. Weiterhin ergibt sich aus den

Regressionen, dass das weibliche Geschlecht in der jüngsten Kohorte keinen signifikanten Einkommenseffekt erwirkt. Im Gegensatz dazu ist der Koeffizient in den beiden älteren Gruppen deutlich größer und signifikant. Der Einfluss des Familienstandes ändert sich je nach Altersgruppe. Zwar ist der Koeffizient in allen Fällen positiv, doch nimmt er mit zunehmendem Alter ab, bis hin zur Insignifikanz für die 51–60-Jährigen. Gleiches nur mit umgekehrten Vorzeichen ist für verheiratete Frauen zu ermitteln.

Kohorten	95 %-Konfidenzintervall
30–40 Jahre	0,074–0,117
41–50 Jahre	0,089 –0,122
51–60 Jahre	0,107–0,140

Tabelle 5.18: 95 %- Konfidenzintervalle für die Bildungsrenditen pro Kohorte; Quelle: DIW, eigene Berechnungen.

Der Einfluss der Unternehmensgröße ist unabhängig vom Alter sehr deutlich. Für die 30–40-Jährigen ist er geringer als in den älteren Kohorten. Hierbei sind die Effekte in der Höhe ähnlich. Der Einfluss eines ländlichen Wohnorts kann für die junge Kohorte nicht signifikant nachgewiesen werden. Der negative Effekt ist erst bei den beiden höheren Alterssegmenten zu verzeichnen. Pendlerverhalten wirkt in allen Altersklassen signifikant positiv auf das Einkommen.

Durch die Separierung der Gruppen können zwischen 31 % und 38 % der Gesamtstreuung mit den Modellen pro Kohorte erklärt werden. Das Modell passt dabei am besten zur mittleren Alterskohorte. Der Wald-Test bestätigt ein weiteres Mal den Gesamterklärungszusammenhang für die gewählte Modellspezifikation.

5.3.2 Öffentlicher Dienst versus Privatwirtschaft

Der Mittelwertvergleich in Abschnitt 4.2 zeigt einen höheren Bruttostundenlohn für Arbeitnehmer des öffentlichen Dienstes. Der Unterschied der mittleren Bruttostundenlöhne aus dem gegebenen Datensatz beläuft sich auf ca. 0,50 € pro Stunde. Da im öffentlichen Sektor anteilig mehr höher qualifizierte Arbeitnehmer beschäftigt sind, ist der mittlere Bruttostundenlohn insgesamt höher. Aus diesem Grund sind in Abschnitt 4.2.3 die Mittelwerte zusätzlich nach ausgewählten Bildungsniveaus aufgeschlüsselt. Aus dem Vergleich ergibt sich eine unterschiedliche Struktur in beiden Sektoren. Fraglich ist nun, ob sich diese verschiedenen Strukturen im öffentlichen und privaten Sektor auch auf die jeweiligen

Bildungsrenditen niederschlagen. Intuitiv ist eine höhere Rendite in der Privatwirtschaft zu erwarten.

Bildungs- gleichung	öffentlicher Dienst	Privat- wirtschaft	Lohn- gleichung	öffentlicher Dienst	Privat- wirtschaft
mittlere väterl. Bild.	2,201*** (0,134)	1,403*** (0,085)	Bildungs- jahre	0,083*** (0,005)	0,100*** (0,005)
hohe väterl. Bildung	4,098*** (0,238)	4,479*** (0,172)	Erfahrung	0,030*** (0,005)	0,029*** (0,005)
Anzahl der Bücher	1,528*** (0,131)	1,436*** (0,084)	Erfahrung ²	-0,000*** (0,000)	-0,000*** (0,000)
Anz. der Schwestern	-0,157** (0,062)	-0,189*** (0,036)	weiblich	-0,040 (0,032)	-0,159*** (0,026)
Anzahl der Brüder	-0,226*** (0,062)	-0,195*** (0,035)	verheiratet	0,086*** (0,030)	0,136*** (0,022)
Staatsan.. nicht deutsch	-1,020*** (0,370)	-1,028*** (0,150)	IAT verh. Frau	-0,094** (0,038)	-0,183*** (0,031)
Konstante	13,305*** (0,396)	12,843*** (0,170)	großes Unternehmen	0,064*** (0,018)	0,280*** (0,015)
			ländliches Gebiet	-0,006 (0,018)	-0,077*** (0,015)
			Pendler	0,055*** (0,018)	0,065*** (0,015)
			Konstante	1,159*** (0,121)	1,085*** (0,109)
korr. R ²	0,37	0,34	korr. R ²	0,35	0,36
Wald-Test	0,000	0,000	Wald-Test	0,000	0,000
n	1.428	3.232	n	1.428	3.232

Tabelle 5.19: Ergebnistabelle der 3SLS-Regression für Arbeitnehmer im öffentlichen Dienst und in der Privatwirtschaft; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

Ergebnisdarstellung. Die Ergebnisse aus der 3SLS-Regression sind in Tabelle 5.19 dargestellt. Hierbei werden im linken Teil der Tabelle die Koeffizienten der Reduced-Form-Regression gezeigt. Die rechte Tabellenseite stellt die Schätzer aus der Lohngleichung dar. Bis auf wenige Ausnahmen sind die Koeffizienten signifikant. Die Bildungsren-

5 Schätzmethoden und Ermittlung der nationalen Bildungsrendite

dite zeigt eine niedrigere Verzinsung im öffentlichen Dienst. Sie liegt bei ca. 8,7 %¹⁵⁶. Der private Sektor hingegen vergütet das zusätzliche Bildungsjahr mit im Mittel ca. 10,5 %. Tendenziell kann damit die eingangs aufgestellte Vermutung für den Vergleich der Bildungsrenditen bestätigt werden. Wie bei der Kohortenanalyse ist ein statistischer Test auf die Unterschiedlichkeit der Koeffizienten nicht möglich.

Daher werden auch an dieser Stelle die 95 %-Konfidenzintervalle herangezogen. Diese sind in Tabelle 5.20 dargestellt. Auch für diese Analyse ergibt sich eine Überschneidung der Intervalle, sodass nicht von signifikanten Unterschieden ausgegangen werden kann. Die Überschneidungsbreite der Intervalle ist jedoch recht klein. Die Ergebnisse können daher durchaus als eine Tendenz verstanden werden.

	95 %-Konfidenzintervall
Öffentlicher Dienst	0,073–0,094
Privatwirtschaft	0,089–0,111

Tabelle 5.20: 95 %-Konfidenzintervalle für die Bildungsrendite im öffentlichen und privaten Sektor; Quelle: DIW, eigene Berechnungen.

Die meisten weiteren Koeffizienten bestätigen den vermuteten Zusammenhang und sind signifikant von null verschieden. Besonders erwähnenswert ist jedoch der Einfluss des weiblichen Geschlechts auf das Einkommen. Der sich ergebende Koeffizient für den öffentlichen Dienst ist zwar negativ, aber insignifikant. Der Familienstand hat im öffentlichen Dienst eine deutliche geringere einkommenssteigernde Wirkung als in der Privatwirtschaft. Ähnliches gilt für den Einfluss der verheirateten Frau. Die einkommensmindernde Wirkung ist in der Privatwirtschaft deutlich stärker als im öffentlichen Dienst. Der Einfluss des Wohnorts ist im öffentlichen Dienst quasi nicht existent. Das Pendlerverhalten wirkt, ebenso wie die meisten Variablen zur Erklärung der Bildungsdauer, recht ähnlich im öffentlichen Dienst und in der Privatwirtschaft. Die Koeffizienten der Reduced-Form-Regression sind alle signifikant und entsprechen den Erwartungen bezüglich der Richtung des Zusammenhangs.

Trotz verschiedener Stichprobenumfänge ist der Erklärungsgehalt der Streuung des logarithmierten Bruttostundenlohns an der gesamten Streuung mit 35 % bzw. 36 % für die Lohngleichung sehr ähnlich. Die Reduced-Form-Regression erklärt 37 % der Gesamtstreuung der Bildungsdauer im öffentlichen Dienst und 34 % in der Privatwirtschaft. Die

¹⁵⁶Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs stets wie folgt berechnet:
 $prozentuale\ Bildungsrendite = (e^{\beta_1} - 1) \cdot 100$.

Stichprobenumfänge unterscheiden sich deutlich. So sind mehr als doppelt so viele Arbeitnehmer dieser Stichprobe in der Privatwirtschaft tätig.

5.3.3 Vollzeit- versus Teilzeitbeschäftigte

Die Gründe für Teilzeitbeschäftigungen sind vielfältig. So werden u. a. Arbeitszeiten verkürzt aufgrund von familiären Verpflichtungen, aus Altersgründen oder um sich parallel fortzubilden. In Abschnitt 4.2.3 sind die mittleren Bruttostundenlöhne sowie die mittleren Bildungsdauern pro Untergruppe angegeben. Obwohl sich die Gruppen vom Bildungsniveau nicht unterscheiden, ergibt sich doch ein signifikanter Einkommensunterschied in Höhe von 3 € zugunsten¹⁵⁷ der Vollerwerbstätigen. Aufgrund dieser großen Unterschiede beim mittleren Bruttostundenlohn ist zu vermuten, dass sich auch die Bildungsrenditen unterscheiden.

Ergebnisdarstellung. In Tabelle 5.22 sind die Regressionsergebnisse angegeben. Vollzeitbeschäftigte haben eine mittlere jährliche Verzinsung von Bildung in Höhe von 9,9 %¹⁵⁸. Für Teilzeiterwerbstätige vergütet sich ein zusätzliches Bildungsjahr im Mittel mit 11,9 %. Teilzeitbeschäftigte (weniger als 30 Wochenarbeitsstunden) erzielen demnach eine höhere Rendite als Vollzeitbeschäftigte. Der Vergleich der Renditen erfolgt im Speziellen wieder mit Blick auf die Konfidenzintervalle, da ein statistischer Test aufgrund der verschiedenen Regressionen nicht anwendbar ist. Diese Intervalle sind in Tabelle 5.21 angegeben. Auch für diese beiden Untergruppen kann nicht von signifikant verschiedenen Bildungsrenditen ausgegangen werden. Aber ähnlich wie bei dem Abschnitt zuvor zeigt der Koeffizient eine Tendenz. Es kann daher geschlussfolgert werden, dass tendenziell Arbeitnehmer in Teilzeit eine höhere Rendite erhalten als jene, die mehr als 30 Stunden pro Woche vertraglich vereinbart haben.

	95 %-Konfidenzintervall
Vollzeit	0,085–0,102
Teilzeit	0,095–0,130

Tabelle 5.21: 95 %-Konfidenzintervalle für die Bildungsrendite voll- und teilzeitbeschäftigter Arbeitnehmer; Quelle: DIW, eigene Berechnungen.

¹⁵⁷Zu beachten ist, dass hierbei Bruttostundenlöhne verglichen werden. Netto kann sich durch die Steuerprogression ein anderes Bild ergeben.

¹⁵⁸Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs stets wie folgt berechnet: $\text{prozentuale Bildungsrendite} = (e^{\beta_1} - 1) \cdot 100$.

5 Schätzmethoden und Ermittlung der nationalen Bildungsrendite

Die Koeffizienten der Lohngleichung sind für die Vollzeitbeschäftigten alle signifikant von null verschieden. In der Höhe sind die Koeffizienten mit denen der anderen Regressionen aus den vorangegangenen Abschnitten vergleichbar, ebenso stimmt die Richtung des Zusammenhangs mit den Vermutungen überein. Der Vergleich zwischen den beiden Untergruppen jedoch zeigt deutliche Unterschiede.

Bildungs- dauer	Lohn- gleichung		Lohn- gleichung	Lohn- gleichung	
	Teilzeit	Vollzeit		Teilzeit	Vollzeit
mittlere väterl. Bildung	1,441*** (0,009)	1,826*** (0,084)	Bildungsjahre	0,112*** (0,009)	0,094*** (0,004)
hohe väterl. Bildung	4,227*** (0,307)	4,370*** (0,157)	Erfahrung	0,018* (0,009)	0,033*** (0,004)
Anzahl der Bücher	1,550*** (0,143)	1,554*** (0,082)	Erfahrung ²	-0,000 (0,000)	-0,000*** (0,000)
Anzahl der Schwestern	-0,136** (0,065)	-0,187*** (0,036)	weiblich	-0,074 (0,091)	-0,099*** (0,022)
Anzahl der Brüder	-0,231*** (0,062)	-0,201*** (0,035)	verheiratet	0,064 (0,107)	0,122*** (0,018)
Staatsangeh. nicht deutsch	-1,235*** (0,297)	-1,067*** (0,163)	IAT verh. Frau	0,091 (0,111)	-0,171*** (0,018)
Konstante	13,102*** (0,328)	13,010*** (0,182)	großes Unternehmen	0,218*** (0,026)	0,218*** (0,013)
			ländliches Gebiet	0,005 (0,026)	-0,069*** (0,013)
			Pendler	0,041 (0,025)	0,067*** (0,013)
			Konstante	0,818*** (0,250)	1,046*** (0,087)
korr. R ²	0,33	0,36	korr. R ²	0,27	0,34
Wald-Test	0,000	0,000	Wald-Test	0,000	0,000
n	1.122	3.571	n	1.122	3.571

Tabelle 5.22: Ergebnistabelle der 3SLS-Regression für Voll- und Teilzeitbeschäftigte; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

Der Einfluss der Berufserfahrung auf das Einkommen ist für Teilzeitbeschäftigte nicht signifikant nachweisbar. Gleiches gilt für das Geschlecht bzw. den Familienstand oder die Kombination aus beidem. Es kann im Gegenteil festgestellt werden, dass der Einfluss auch für den in Teilzeit beschäftigten Arbeitnehmer in Unternehmen mit mehr als 200 Mitarbeitern signifikant ist. Das Pendlerverhalten hat dagegen keinen signifikanten Einfluss auf

das Einkommen bei Teilzeiterwerbstätigen. Das generell geringere Einkommensniveau ist im Achsenabschnitt erkennbar.

Die Modellspezifikation kann für die Teilzeitbeschäftigten lediglich 27 % der Gesamtstreuung erklären. Im Vergleich zu den bisher dargestellten Modellen ist dies eine deutlich schlechtere Modellgüte. Der Anteil für die Vollzeitbeschäftigten ist mit 34 % dagegen vergleichbar mit den anderen Ergebnissen. Es muss daher davon ausgegangen werden, dass es für Teilzeitbeschäftigte andere wichtige Einflussfaktoren für das Einkommen gibt, die an dieser Stelle nicht berücksichtigt sind.

Beim Bildungsniveau zeigt sich bereits in der deskriptiven Analyse nur wenig Heterogenität zwischen beiden Gruppen. Ähnliches kann auch durch die Reduced-Form-Regression ermittelt werden. Alle Koeffizienten sind mit einer Irrtumswahrscheinlichkeit von 5 % signifikant von null verschieden und zeigen die erwarteten Vorzeichen. Auch die Ausprägungen für beide sind Subgruppen ähnlich. Das R^2 liegt für Teilzeitbeschäftigte bei 33 %, für Vollzeitbeschäftigte bei 36 %.

5.3.4 Zusammenfassung

Die Analyse beschäftigt sich mit drei Separierungen. Alle Regressionen werden pro Subgruppe mit dem 3SLS-Verfahren durchgeführt und erhalten die gleiche Modellspezifikation. Im ersten Schritt werden drei verschiedene Kohorten betrachtet. Die Bildungsrenditen unterscheiden sich deutlich in der Höhe und steigen mit zunehmendem Alter. Vergleicht man jedoch 95 %-Konfidenzintervalle, so ist diese Unterschiedlichkeit nicht nachweisbar. Die zweite Separierung wird für Arbeitnehmer im öffentlichen Dienst und in der Privatwirtschaft vorgenommen. Dabei wird eine höhere mittlere Verzinsung für Personen in der Privatwirtschaft ermittelt. Doch auch bei dieser Separierung ergibt sich beim Vergleich der Vertrauensbereiche eine Überschneidung der Intervalle. Gleiches konnte auch für die Separierung in Voll- und Teilzeiterwerbstätige ermittelt werden. Zwar zeigt sich eine geringere Rendite für Vollzeiterwerbstätige, diese ist jedoch im Vergleich der Konfidenzintervalle nicht eindeutig abgrenzbar. Demzufolge kann eine Tendenz für eine höhere Rendite für ältere Arbeitnehmer, Mitarbeiter im öffentlichen Dienst sowie Teilzeiterwerbstätige festgestellt werden. Die Unterschiede zwischen den Subgruppen sind jedoch nicht groß genug, um sie als signifikant einzustufen.

5.4 Berücksichtigung der Erwerbslosen

Die bisherige Analyse bezieht sich ausschließlich auf Arbeitnehmer. Alle Nichterwerbstätigen werden vernachlässigt. Es wird angenommen, dass sich die Gruppe der Erwerbstätigen von der Gruppe der Nichterwerbstätigen unterscheidet.¹⁵⁹ Aus dem Datensatz ergibt sich, dass alle bisherigen Aussagen lediglich für die nicht selbstständig beschäftigten Arbeitnehmer zwischen 30 und 60 Jahren getroffen werden können. Durch die zu vermutende Andersartigkeit der Nichterwerbstätigen ist von einem Selbstselektionsproblem durch die Partizipation oder Nichtpartizipation am Arbeitsmarkt auszugehen. Dieses wurde bereits in Abschnitt 5.2 kurz aufgezeigt und ist Thema dieses Abschnitts. Zunächst wird dabei dieses Problem theoretisch fundiert. Im nächsten Schritt sind zwei Möglichkeiten zur Berücksichtigung der Erwerbslosen diskutiert. Dabei erfolgt im ersten Ansatz eine Erweiterung des Datensatzes um ALG-I-Empfänger. Im zweiten Schritt werden mit Verwendung des Heckman-Verfahrens alle Erwerbspersonen berücksichtigt. Hierbei werden zunächst Unterschiede zwischen den Arbeitenden und Nichtarbeitenden identifiziert. Diese werden im zweiten Schritt als Kontrollgröße in der Lohnleichung berücksichtigt.

5.4.1 Theoretische Fundierung

Gronau (1974) entwickelt eine theoretische Formulierung für das beschriebene Problem. Die Darstellung des Modells basiert auf den Ausführungen von Wooldridge (2002), S. 561 ff. Ziel dieses Modells ist die Diskussion über die aktive Arbeitsmarkteteiligung einer Erwerbsperson. Ausgangspunkt ist dabei ein Modell welches den individuellen Nutzen maximiert. Dabei hat das Individuum die wöchentlichen Arbeitsstunden als Aktionsparameter. Es gilt,

$$\max_{h_i} U(w_i^o h_i + a_i, h_i), \quad (5.1)$$

wobei

$$0 \leq h_i \leq h_i^{\max}$$

ist. Der Nutzen U ergibt sich in dieser Funktion einerseits aus dem Einkommen, andererseits aus dem Arbeitsleid. Einkommen q wird unterschieden in Arbeits- und Nichtar-

¹⁵⁹Die genaue Prüfung folgt in diesem Abschnitt.

beitseinkommen. Ersteres setzt sich zusammen aus dem Stundenlohn w_i^o multipliziert mit der Anzahl der geleisteten Arbeitsstunden pro Woche h_i . Das Nichtarbeitseinkommen wird beschrieben mit a_i . Jegliches Einkommen wirkt bei Zunahme nutzensteigernd. Gegenteilig wirkt das Arbeitsleid in der Nutzenfunktion. Dieses, gemessen in der Zahl der geleisteten Arbeitsstunden, hat bei Zunahme eine nutzenmindernde Wirkung. Bei Betrachtung der Extrempunkte kann zum einen überhaupt nicht gearbeitet werden, dann ist $h_i^{min} = 0$. Maximal können zum anderen in einer Woche h_i^{max} zur Generierung von Arbeitseinkommen genutzt werden. Die Arbeitszeit wirkt nun in zwei Richtungen auf den Nutzen: positiv durch ein höheres Einkommen und negativ durch das entstehende Arbeitsleid. Es gilt, wenn

$$\frac{dU}{dh_i} < 0, \quad (5.2)$$

so wird die gearbeitete Stundenzahl pro Woche nicht erhöht. Im Speziellen, wenn das Individuum i an der Stelle $h_i = 0$ bereits keinen positiven Grenznutzen aus einer zusätzlichen Arbeitsstunde erzielt, wird es auf dem Arbeitsmarkt gar nicht erst aktiv. Dies mag an großem Arbeitsleid bzw. einer geringen Entlohnung liegen. Aus dem Maximierungskalkül hinsichtlich der Frage der generellen Aktivität am Arbeitsmarkt lässt sich ein theoretischer Reservationslohn w^r ermitteln. Für ihn gilt, dass

$$w_i^r = -\frac{GN_{h_i}(a_i, 0)}{GN_{q_i}(a_i, 0)} > 0. \quad (5.3)$$

Da eine zusätzliche Arbeitsstunde negativ auf den Nutzen wirkt, siehe Gleichung 5.2, ist der Zähler negativ. Der Grenznutzen GN aus Einkommen ist positiv. Demnach ergibt sich für den Reservationslohn ein strikt positiver Wert. Das Individuum wird sich am Arbeitsmarkt beteiligen, wenn gilt, dass

$$w_i^o \geq w_i^r, \quad (5.4)$$

andernfalls nicht. Es muss davon ausgegangen werden, dass sich die Personen, bei denen die Ungleichung erfüllt ist, von den Individuen, für die der Reservationslohn größer ist als das potentielle Arbeitseinkommen, unterscheiden. Daher wird nachfolgend eine Analyse für die verschiedenen Personengruppen vorgenommen.

5.4.2 Bildungsrenditen für Erwerbstätige und ALG-I-Empfänger

Um dem Unterschied gerecht zu werden, gilt es in diesem Abschnitt, den Datensatz um die nichterwerbstätigen Individuen zu erweitern. Hierfür folgt zunächst eine Diskussion über die Möglichkeiten zur Berücksichtigung Erwerbsloser in der Stichprobe, um dann eine Schätzung vorzunehmen.

Definition der Erwerbslosen und Bemessungsgrundlage der Einnahmen. Zur Verallgemeinerung der Aussagen wird der Datensatz erweitert. Dabei gilt es bei der Berechnung der Bildungsrenditen jene mit zu berücksichtigen, die im Jahr 2008 arbeitslos waren. Als arbeitslos gilt jemand, der nach § 119 SGB III¹⁶⁰

- „nicht in einem Beschäftigungsverhältnis steht (Beschäftigungslosigkeit),
- sich bemüht, seine Beschäftigungslosigkeit zu beenden (Eigenbemühungen) und
- den Vermittlungsbemühungen der Agentur für Arbeit zur Verfügung steht (Verfügbarkeit).“

Diejenigen haben Anspruch auf staatliche Transferzahlungen. Nun wird unterschieden zwischen Arbeitslosengeld¹⁶¹ I und der Grundsicherung für Arbeitssuchende, ALG II. Gemäß § 117 SGB III haben nur die Personen Anspruch auf ALG I, die arbeitslos sind oder sich in einer Weiterbildungsmaßnahme befinden und noch nicht Rentner sind. Im § 127 SGB III wird die Anspruchsdauer geregelt. Dabei wird der Anspruchszeitraum bestimmt durch die Dauer des zuvor bestandenen Versicherungspflichtverhältnisses und dem Lebensalter. Generell kann dieser zwischen drei und 24 Monaten liegen. In diesem Zeitraum wird anteilig zum alten Nettogehalt ein allgemeiner oder erhöhter Leistungssatz ausgezahlt. Der erhöhte Leistungssatz wird gültig, wenn mindestens ein Kind¹⁶² dem Haushalt zugehörig ist. Der allgemeine Leistungssatz beträgt 60 %, der erhöhte 67 % des ursprünglichen Gehalts.

Die Bemessungsgrundlage für ALG II erfolgt mit detaillierter Beschreibung im SGB II. Sie hängt von diversen Faktoren ab. Hierzu zählen unter anderem das eigene Vermögen sowie mögliche Unterstützungsleistungen durch den Lebenspartner. Sie dient gemäß § 1

¹⁶⁰SGB ist das Sozialgesetzbuch.

¹⁶¹Im folgenden kurz als ALG bezeichnet.

¹⁶²Das Kind muss im Sinne des § 32 Abs. 1, 3 und 5 EStG dem Arbeitslosen oder dessen Ehepartner zugehörig sein.

Abs. 1 SGB II dazu, den „Leistungsberechtigten [...] ein Leben zu [ermöglichen] [...], das der Würde des Menschen entspricht.“

Erweiterter Datensatz (ALG-I-Empfänger) und deskriptive Statistik. In diesem Abschnitt wird zunächst die sinnvolle Berücksichtigung der ALG-I- und/ oder ALG-II-Empfänger diskutiert. Darauf aufbauend wird der Datensatz erweitert und deskriptive Statistiken werden aufgezeigt.

ALG I. Ein Empfänger von ALG I erhält monatlich einen bestimmten Anteil des vormaligen Gehalts aus der vorhergehenden Beschäftigung. Die Höhe dieses früheren Einkommens ist durch ähnliche Einflüsse bestimmt wie das der derzeit Beschäftigten. Demnach ist es sinnvoll, den Datensatz um diese Gruppe zu erweitern und eine Schätzung auf die Einnahmen durchzuführen.

ALG II. Im Gegensatz zu ALG-I-Empfängern erhält ein Empfänger von ALG II einen generell definierten absoluten Betrag. Der Maximalbetrag im Jahr 2008 lag unabhängig vom Bildungsniveau bei 347 €. Bei Berücksichtigung dieser Gruppe stellt sich die Frage, welcher Betrag für die abhängige Variable – die Einnahmen – verwendet wird. Es könnte zum einen das Arbeitseinkommen der ALG-II-Empfänger betrachtet werden. Dieses ist null. Schätzt man die Bildungsrenditen auf Basis des Mincer-Modells, wie beschrieben in Abschnitt 5, so ergäbe sich, dass

- die Zahl null als eine tatsächliche Ausprägung des Einkommens anzusehen wäre. Tatsächlich jedoch handelt es sich lediglich um einen unbeobachtbaren Wert. Wie in Abschnitt 5.4.1 gezeigt, ist davon auszugehen, dass für diese Gruppe der Reservationslohn größer ist als das potentielle Einkommen aus Arbeit. Wooldridge (2002), S. 562, schreibt dazu „Setting y_1 to zero (or any value) when $y_2 = 0$ is misleading and can lead to inappropriate use of the model. For example, it makes no sense to set the wage offer to zero just because we do not observe it.“
- die Zusammensetzung vom Einkommen („null“) sich nicht anhand der zuvor identifizierten Determinanten wie Bildung, Berufserfahrung und andere Faktoren erklären lässt. Die ursprünglich identifizierte Schätzgleichung kann demnach nicht verwendet werden.

- die solidarische Grundsicherung nicht als Substitut zum Gehalt berücksichtigt ist. Die Einnahmen sind nicht null, lediglich jene aus der Arbeitstätigkeit sind nicht vorhanden.

Aufgrund dieser Konstellation können ALG-II-Empfänger im Datensatz für das Arbeitseinkommen nicht die Ausprägung „null“ erhalten. Weiterhin möglich wäre, die Höhe des monatlichen ALG-II-Betrages zu verwenden. Hiergegen sprechen jedoch drei Argumente:

- Die Höhe der Grundsicherung ergibt sich durch diverse Faktoren, zum Beispiel aus dem Einkommen des Partners. Demnach ist die Berücksichtigung der Höhe der ausbezahlten Grundsicherung ohne Hinzunahme diverser weiterer Lebensumstände willkürlich.
- Die Höhe des erhaltenen Sicherungsbetrages wird in keiner Weise durch das Bildungsniveau bestimmt.
- Der ausbezahlte Betrag berücksichtigt ausschließlich die Grundsicherung der Lebenshaltung, exklusive Mietkosten und Ähnlichem. Eine Vergleichbarkeit der Beträge zwischen Erwerbstätigen und ALG-II-Empfängern ist nicht gegeben.

Für die weitere Betrachtung bedeutet das, dass die Analyse auf Empfänger von ALG I ausgeweitet wird und damit das Risiko der Erwerbslosigkeit zumindest teilweise mit Berücksichtigung findet. Die Hinzunahme von ALG-II-Empfängern hingegen zeigt sich als inhaltlich und technisch ungeeignet.

Erweiterung: Erwerbstätige und ALG-I-Empfänger. Die Datensatzreduzierung aus der Grunddatei des SOEP ist bereits in Abschnitt 4.1 beschrieben. Zu den 4.693 bisher betrachteten Erwerbstätigen werden 119 ALG-I-Empfänger hinzugezogen. Insgesamt werden demnach 4.812 Erwerbstätige und ALG-I-Empfänger zwischen dem 30. und 60. Lebensjahr untersucht, die sich nicht im Ausbildungsprozess befinden.

Deskriptive Statistik. Die Gruppe der ALG-I-Empfänger ist lediglich mit einem Anteil von 2,5 % in der Stichprobe vertreten. Es ist zu erwarten, dass die Hinzunahme dieser Gruppe sich nur marginal auf die gesamtdeutsche Bildungsrenditen auswirkt. Weitergehende Untersuchungen durch Bildung von Untergruppen versprechen nur geringe Veränderungen in den bereits geschätzten Resultaten. Daher wird ausschließlich eine Untersuchung für das gesamte Bundesgebiet durchgeführt.

Zur Beschreibung des Datensatzes wird der um die ALG-I-Empfänger erweiterte Datensatz mit dem ursprünglichen Datensatz ohne Arbeitslose und dem separierten Datensatz für die 119 ALG-I-Empfänger verglichen. In Tabelle 5.23 sind hierfür ausgewählte deskriptive Statistiken präsentiert. Für den Vergleich zwischen erweitertem und ursprünglichem Datensatz mit ausschließlicher Berücksichtigung der Erwerbstätigen kann für die Anteile erwartungsgemäß kein Unterschied festgestellt werden. Betrachtet man die ALG-I-Empfänger separat, so ist der Frauenanteil mit 48 % in ähnlicher Höhe. Eine unterschiedliche Verteilungsstruktur ist bei der Separierung nach Kohorten zu bemerken. Die älteste Kohorte ist bei Betrachtung der Arbeitslosen am meisten vertreten.

Anteil	Erweiterter Datensatz	Erwerbstätige	ALG-I-Empfänger
weiblich	0,47	0,47	0,48
30–40 Jahre	0,30	0,30	0,32
41–50 Jahre	0,40	0,40	0,29
50–60 Jahre	0,30	0,30	0,39
n	4.812	4.693	119

Tabelle 5.23: Deskriptive Statistiken des erweiterten Datensatzes mit ALG-I-Empfängern gesamt und separat; Quelle: DIW, eigene Berechnungen.

Für den Vergleich von Mittelwerten im Rahmen der deskriptiven Statistik werden die mittlere Bildungsdauer sowie der mittlere Stundenlohn verglichen. Es ergibt sich, dass sich die durchschnittlichen Bildungsdauern in beiden Datensätzen nicht wesentlich unterscheiden. Die ALG-I-Empfänger separat betrachtet ergeben dagegen wiederum ein abweichendes Bild. So haben diese eine mittlere Bildungsdauer von 11,6 Jahren.

	Erweiterter Datensatz	Erwerbstätige	ALG-I-Empfänger
Durchschnittliche Bildungsdauer in Jahren	12,7 (2,747)	12,8 (2,754)	11,6 (2,181)
Mittlerer Nettomonatslohn in €	1.729,74 (998,3)	1.752,97 (997,3)	813,70 (472,0)
n	4.812	4.693	119

Tabelle 5.24: Mittelwertvergleich des erweiterten Datensatzes mit ALG-I-Empfängern gesamt und separat; Quelle: DIW, eigene Berechnungen.

Der Auszahlungsbetrag der ALG-I-Empfänger ist eine Nettogröße. Um daher eine einheitliche Basis zu schaffen, wird auch der Nettolohn für die Erwerbstätigen herangezogen. Eine Berechnung des Stundenlohns ist aufgrund mangelnder Kenntnis über die zuvor geleisteten Arbeitsstunden nicht möglich. Demnach wird im Modell als abhängige Variable der Nettomonatslohn für beide Datensätze einbezogen.

Durch die Hinzunahme der ALG-I-Empfänger wird erwartungsgemäß der mittlere Nettolohn im Vergleich leicht verringert. Betrachtet man die ALG-I-Empfänger separat, so ergeben sich lediglich Einnahmen in Höhe von 813,70 €. Dieser Unterschied lässt sich nicht nur auf die Verringerung des Einkommens auf 60 % bzw. 67 % des vorherigen Niveaus zurückführen. Durch die geringe Bildungsdauer ist hier von einem selektierenden Verhalten auf die beiden Erwerbsgruppen auszugehen.

Modellbeschreibung und Ergebnisdarstellung. Die Überprüfung des Einflusses der ALG-I-Empfänger erfolgt weiterhin über die 3SLS-Methode.¹⁶³ In Abbildung 5.6 wird die Vorgehensweise für den Vergleich der gesamtdeutschen Bildungsrenditen auf Basis eines Datensatzes mit Erwerbstätigen und ALG-I-Empfängern dargestellt.

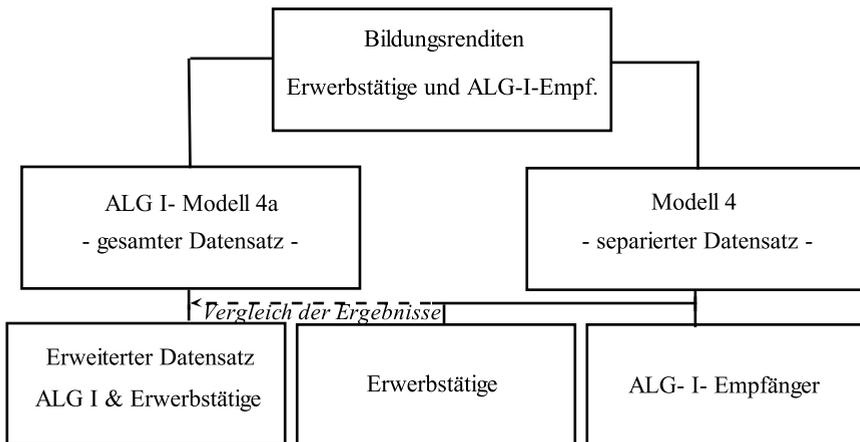


Abbildung 5.6: Überblick zur Schätzung der Bildungsrenditen für Erwerbstätige und ALG-I-Empfänger, eigene Darstellung.

¹⁶³Die nähere Beschreibung findet sich in Abschnitt 5.2.2.

Im ersten Schritt wird dabei die Bildungsrendite für den erweiterten Datensatz berechnet. Hierfür wird die ursprüngliche Lohngleichung¹⁶⁴ um eine Dummy-Variable für die Arbeitslosen erweitert. Durch die Verwendung des logarithmierten Nettolohns ist die sich ergebende Bildungsrendite nicht eindeutig vergleichbar mit dem ursprünglich¹⁶⁵ gesamtdeutschen Ergebnis. Um dies herzustellen, wird die Regression für den ursprünglichen Datensatz auf den Nettomonatslohn nochmals durchgeführt. Im nächsten Schritt wird der erweiterte Datensatz nach dem Erwerbsstatus separiert und jeweilig pro Gruppe geschätzt. Dabei kann auch eine Bildungsrendite, die ausschließlich für ALG-I-Empfänger gilt, ermittelt werden.

ALG I – Modell 4a: Erweiterter Datensatz. Das in Abschnitt 5.1.1 beschriebene Modell 4¹⁶⁶ wird leicht modifiziert. Dabei werden zwei Anpassungen vorgenommen. Auf der einen Seite müssen Variablen aufgrund fehlender Ausprägungen für die ALG-I-Empfänger weggelassen werden und zum anderen muss das generell niedrigere Einkommensniveau dieser Gruppe berücksichtigt werden. Betrachtet man beispielsweise die Dummy-Variable für die Unternehmensgröße, so gibt es hierbei genau zwei Ausprägungen. Eine genaue Zuordnung zu einer dieser Ausprägungen ist für die ALG-I-Empfänger nicht möglich, da ihnen weder eine null – arbeitend in einem kleinen Unternehmen – noch eine 1 – arbeitend in einem großen Unternehmen – zugeordnet werden kann. Daher findet eine leichte Modifikation des Modells statt. Zusätzlich entspricht das Arbeitslosengeld generell lediglich 60 % bzw. 67 % des ursprünglichen Einkommens, sodass das generell geringere Niveau sich nicht verzerrend auf die Bildungsrendite auswirken darf. Daher wird eine Dummy-Variable für die Arbeitslosigkeit hinzugenommen. Das Regressionsmodell für die Lohngleichung lautet nun:

$$\begin{aligned} \ln Y_i^{\text{netto}} &= \beta_0 + \beta_1 S_i + \beta_2 \text{Er} f_i + \beta_3 \text{Er} f_i^2 + \beta_4 \text{weibl}_i + \beta_5 \text{verh}_i \\ &+ \beta_6 \text{IAT}_{W-\text{verh}_i} + \beta_7 D_{AL_i} + \varepsilon_i, \end{aligned}$$

wobei der Korrekturterm in Tabelle 5.25 näher dargestellt ist. Im Unterschied zum vorangegangenen Modell wird nun der Nettomonatslohn verwendet.

¹⁶⁴Vgl. Abschnitt 5.1.1.

¹⁶⁵Vgl. Abschnitt 5.1.2.

¹⁶⁶Die Bezeichnung ALG I-Modell 4a ergibt sich aufgrund der lediglich geringfügigen Modifizierung des Modells aus Abschnitt 5.

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
D_{AL_i}	Arbeitslos	dichotom (1: Empfänger von ALG-I, 0: Erwerbstätige)	negativ

Tabelle 5.25: Variablenbezeichnung – Erweiterung um ALG-I-Empfänger.

Als eine sinnvolle Methode für die Schätzung von Bildungsrenditen ist das 3SLS-Verfahren identifiziert worden. Daher wird auch für den erweiterten Datensatz die Reduced-Form-Regression, d. h. die Schätzung auf die Bildungsdauer, durchgeführt. Die Diskussion für die beste Modellspezifikation ist in Abschnitt 5.2.1 bereits gezeigt. Aufbauend darauf wird weiterhin von dieser Gleichung für die Bildungsdauer ausgegangen.

Ergebnisdarstellung. In Tabelle 5.26 sind die verschiedenen Ergebnisse je nach Regressionsmodell dargestellt. Spalte 1 zeigt die Ergebnisse für den erweiterten Datensatz mit Berücksichtigung der Erwerbstätigen und ALG-I-Empfänger. Hierbei ist eine Korrektur um die verschiedenen Einkommensniveaus zwischen beiden Gruppen notwendig. Dafür wird zusätzlich die Dummy-Variable hinzugezogen. Die folgenden beiden Spalten zeigen dann die Ergebnisse aus dem separierten Datensatz, wobei zunächst lediglich die Erwerbstätigen betrachtet, in Spalte 3 nur die ALG-I-Empfänger analysiert werden. Dabei unterscheiden sich die Ergebnisse der Spalte 2 lediglich durch die Verwendung des Nettolohns und die kleine Variablenreduktion im Regressionsmodell von den in Abschnitt 5.2.2 diskutierten. Für den erweiterten und ursprünglichen Datensatz kann in diesen Spezifikationen eine Verzinsung des zusätzlichen Bildungsjahres im Mittel in Höhe¹⁶⁷ von 10,2 % ermittelt werden. Die zusätzliche Berücksichtigung der ALG-I-Empfänger im erweiterten Datensatz hat demnach keinen Einfluss auf die Höhe der Bildungsrendite. Auch die Verwendung des Nettomonatslohns im Vergleich zum Bruttostundenlohn¹⁶⁸ führt zu keiner Veränderung des Ergebnisses.

Anders hingegen ist es, wenn die Bildungsrendite separat für ALG-I-Empfänger berechnet wird. Diese erhalten im Mittel lediglich eine Verzinsung von 8,7 %. Aufgrund der verschiedenen Kovarianzmatrizen sind auch diese separat berechneten Bildungsrenditen nicht über einen statistischen Test miteinander vergleichbar.

¹⁶⁷Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs wie folgt berechnet:
 $prozentuale\ Bildungsrendite = (e^{\beta_1} - 1) \cdot 100.$

¹⁶⁸Vgl. Regressionsergebnisse in Abschnitt 5.2.2.

Lohngleichung (Nettolohn)	erweiterter Datensatz	Erwerbstätige	ALG-I- Empfänger
	ALG I – Modell 4a	Modell 4	Modell 4
Bildungsjahre	0,097*** (0,004)	0,097*** (0,004)	0,083** (0,035)
Erfahrung	0,024*** (0,004)	0,024*** (0,004)	0,008 (0,027)
Erfahrung ²	-0,000*** (0,000)	-0,000*** (0,000)	-0,000 (0,000)
weiblich	-0,256*** (0,023)	-0,257*** (0,024)	-0,325** (0,142)
verheiratet	0,212*** (0,020)	0,206*** (0,021)	0,373*** (0,120)
IAT verh. Frau	-0,460*** (0,028)	-0,460*** (0,023)	-0,370** (0,180)
Arbeitslos	-0,681*** (0,042)		
Konstante	5,794*** (0,085)	5,788*** (0,086)	5,527*** (0,517)
R^2	0,43	0,41	0,36
Wald-Test	0,000	0,000	0,000
n	4.812	4.693	119

Tabelle 5.26: Ergebnistabelle der Lohngleichung der 3SLS-Regression für den erweiterten (Erwerbstätig und ALG I) Datensatz zusammen und getrennt sowie den ursprünglichen Datensatz im Vergleich; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

Aufgrund des geringen Stichprobenumfangs von lediglich 119 Beobachtungen bei den ALG-I-Empfängern empfiehlt es sich, die 95 %-Konfidenzintervalle als Vergleichsgröße heranzuziehen, da davon auszugehen ist, dass der kleine Stichprobenumfang einen deutlich größeren Spielraum zulässt, als der der Erwerbstätigen.

In Tabelle 5.27 werden diese dargestellt. Der Vergleich¹⁶⁹ bestätigt die Vermutung und zeigt demnach eine deutliche Überlagerung der Intervalle. Daher kann trotz des, in der Höhe geringeren Koeffizienten kein signifikanter Unterschied in der Höhe der Bildungsrenditen festgestellt werden.

¹⁶⁹Der Vergleich wird beschränkt auf die separierte Betrachtung, da die Darstellung des gemeinschaftlichen Datensatzes lediglich eine Kombination aus beiden Datensätzen ist.

5 Schätzmethoden und Ermittlung der nationalen Bildungsrendite

	Bildungsrendite	95 %-Konfidenzintervall
Erwerbstätige	0,097	0,088–0,105
ALG-I-Empfänger	0,083	0,015–0,152

Tabelle 5.27: 95 %-Konfidenzintervalle im Vergleich: separierter Datensatz; Quelle: DIW, eigene Berechnungen.

Auch der Vergleich der weiteren Koeffizienten der Lohngleichung zeigt in der Höhe und Ausrichtung ähnliche Koeffizienten. Die geringen Abweichungen bei der ALG-I-Separierung lassen sich durch den geringen Stichprobenumfang begründen. Die Vorzeichen entsprechen bei den drei Regressionen den Erwartungen. In der zusammengefassten Regression zeigt auch der Korrekturfaktor den erwarteten negativen Zusammenhang auf das Einkommen. Die Koeffizienten der Reduced-Form-Regression sind im Anhang A.3.6 in Tabelle A.16 dargestellt. Das Bestimmtheitsmaß ist mit 36 % bis 43 % erklärter Streuung an der Gesamtstreuung für alle Regressionen zufriedenstellend. Ebenso kann ein weiteres Mal die Gesamtsignifikanz durch den Wald-Test bestätigt werden.

Zusammenfassung

Die Schätzung der Lohngleichung zur Ermittlung der Bildungsrendite ergibt für alle drei Modelle in der Höhe ähnliche Bildungsrenditen. Es kann durch die Berücksichtigung der ALG-I-Empfänger kein Unterschied identifiziert werden, sodass von einer Rendite in Höhe von ca. 10,2 % auszugehen ist. Es muss jedoch festgestellt werden, dass das Einbeziehen aller Arbeitslosen technisch schwierig ist und demnach lediglich ALG-I-Empfänger untersucht werden. Für diese Gruppe jedoch zeigt sich aufgrund der geringen Beobachtungszahl nur ein geringer Einfluss, wenn die Bildungsrenditen für Erwerbstätige und ALG-I-Empfänger zusammen berechnet werden. Betrachtet man dagegen die Gruppen separat, ist zwar eine signifikante Rendite für die Erwerbslosen schätzbar, das 95 %-Konfidenzintervall hat hingegen aufgrund des geringen Stichprobenumfangs eine große Spannweite, sodass Unterschiede zu den Erwerbstätigen nicht feststellbar sind. Eine sinnvolle aussagekräftige Analyse mit Arbeitslosen ist durch das Hinzufügen der ALG-I-Empfänger nur schwer möglich. Beim Vergleich der deskriptiven Statistiken ergeben sich jedoch Unterschiede in der mittleren Bildungsdauer. Es ist nicht auszuschließen, dass sich die Gruppe der Arbeitslosen von jenen der Erwerbstätigen unterscheidet. Insbesondere muss dabei die bisher unberücksichtigte Gruppe der ALG-II-Empfänger mit einbezogen werden.

5.4.3 Bildungsrenditen für alle Erwerbspersonen

Ziel dieses Abschnitts ist die Berücksichtigung aller Erwerbslosen. Zuvor wurde gezeigt, dass die Ausweitung der Analyse durch die einfache Erweiterung des Datensatzes nur teilweise realisierbar ist. Daher gilt es, einen Korrekturterm zu entwickeln, der für die Unterschiede zwischen Erwerbstätigen und Erwerbslosen kontrolliert, so dass dies in der Berechnung der Bildungsrendite nicht verzerrend wirkt. Hierfür wird zunächst, die in Abschnitt 5.4.1 beschriebene, die theoretische Fundierung erweitert, mit dem Ziel der Ermittlung zweier Schätzgleichungen. Das Verfahren von Heckman¹⁷⁰ schätzt zunächst die Arbeitsmarktbeteiligung, um im zweiten Schritt die zu schätzende Lohngleichung um einen Korrekturfaktor zu erweitern, der aus der ersten Regression bestimmt wurde. Damit kann für die Unterschiedlichkeit zwischen Arbeitslosen und Erwerbstätigen kontrolliert werden. Nach der theoretischen Diskussion wird das Modell für den gesamtdeutschen Raum geschätzt und über den regionalen Vergleich ausgebaut.

Theoretische Fundierung der Zwei-Schritt-Methode von Heckman. Durch die Nichtpartizipation am Arbeitsmarkt sind für Erwerbslose keine Beobachtungen für das Einkommen erhältlich. Als Pendant zum Erwerbseinkommen gilt für sie der Reservationslohn. Dieser ist empirisch nicht beobachtbar. Für Erwerbslose kann lediglich festgehalten werden, dass der Reservationslohn höher ist als der potentielle Lohn auf dem Arbeitsmarkt.¹⁷¹

Aufbauend auf die Darstellung in Abschnitt 5.4.1 können zwei Schätzgleichungen identifiziert werden. Zum einen jene zur Erklärung des beobachtbaren Arbeitseinkommens, zum anderen jene zur Erklärung des unbeobachtbaren Reservationslohns. Dabei ist die Schätzbarkeit der zweiten Gleichung nicht gegeben. Allgemein gilt jedoch, dass

$$\log w_i^o = \beta_1 X_{i1} + u_{i1} \quad (5.5)$$

$$\log w_i^r = \beta_2 X_{i2} + \gamma_2 a_i + u_{i2}, \quad (5.6)$$

wobei die Fehlerterme (u_{i1}, u_{i2}) unabhängig von den erklärenden Variablen (X_{i1}, X_{i2}, a_i) sind. Das Arbeitseinkommen w_i^o wird hierbei erklärt durch die Eigenschaf-

¹⁷⁰Diese Ausführungen folgen Gronau (1974), Heckman (1979) und Wooldridge (2002), S. 562 ff.

¹⁷¹Vgl. Gronau (1974).

ten X_{i1} ; der Reservationslohn durch X_{i2} , zuzüglich des Einflusses des Nichtarbeitseinkommens. Die Erwerbperson wird nur dann am Arbeitsmarkt aktiv, wenn der Reservationslohn zumindest dem gezahlten Arbeitseinkommen entspricht. Berechnet man die theoretische Differenz aus beiden Größen, so ergibt sich aus der Kombination von Gleichungen (5.5) und Gleichung (5.6) folgender Ausdruck:

$$\begin{aligned} \log w_i^o - \log w_i^r &= \beta_1 X_{i1} - \beta_2 X_{i2} - \gamma_2 a_i + u_{i1} - u_{i2} \\ &\equiv \delta_2 X_i + v_{i2} \geq 0. \end{aligned} \quad (5.7)$$

Dabei gilt, dass $\beta_1 X_{i1} - \beta_2 X_{i2} = \delta_2 X_i$ und $u_{i1} - \gamma_2 a_i - u_{i2} = v_{i2}$. Werden ausschließlich Erwerbstätige berücksichtigt, ergibt sich ein Selektionsproblem in Höhe des Ausdrucks in Gleichung 5.7, da für alle Erwerbstätigen die Ungleichung erfüllt ist.

Mit den bisher dargestellten Verfahren ist die Schätzung einer Bildungsrendite ohne die verzerrende Wirkung durch die Selbstselektion nicht¹⁷² möglich. Demnach muss ein anderer Ansatz gewählt werden. Hierfür kann nun formuliert werden, dass

$$y_{i1} = \log w_i^o, \quad (5.8)$$

$$y_{i2} = \begin{cases} 1, & \text{wenn } (\delta_2 X_i + v_{i2} \geq 0) \\ 0, & \text{sonst} \end{cases} \quad (5.9)$$

Es gilt, dass das Arbeitseinkommen nur dann beobachtbar ist, wenn die dichotome Variable y_{i2} die Ausprägung 1 hat und demnach die Ungleichung 5.7 erfüllt ist. In diesem Fall ist das Arbeitseinkommen größer oder gleich dem Reservationslohn des Individuums. Erweitert man Gleichung (5.8) um Gleichung (5.5), so ergibt sich

$$y_{i1} = \log w_i^o = \beta_1 X_{i1} + u_{i1}.$$

¹⁷²Vgl. Abschnitt 5.4.2.

Im Gegensatz zu Gleichung (5.8), die nicht für alle Erwerbspersonen beobachtbar ist, kann für Gleichung (5.9) für alle Individuen eine Ausprägung zugeordnet werden. Ist die Person erwerbstätig, so nimmt y_{i2} den Wert 1 an, andernfalls 0. Weiterhin gilt die übliche Annahme, dass u_{i1} und v_{i2} unabhängig von X_i sind. Die Differenz der Residuen, $v_{i2} = u_{i1} - u_{i2}$, sei standardnormalverteilt. Für den Erwartungswert des Residuums wird die Beziehung $E[u_{i1} | v_{i2}] = \gamma_{i1} v_{i2}$ angenommen.

Das Ziel der Regression ist die Schätzung einer allgemeinen mittleren prozentualen Verzinsung von Humankapital, unabhängig von den Eigenschaften der Individuen. Daher muss für die Eigenschaften, die zur Selektion führen, in der Lohngleichung kontrolliert werden. Allgemein kann damit formuliert werden, dass

$$\begin{aligned} E[y_{i1} | X_i, v_{i2}] &= \beta_1 X_{i1} + E[u_{i1} | X_i, v_{i2}] \\ &= \beta_1 X_{i1} + \gamma_{i1} v_{i2}. \end{aligned}$$

Ist der Fehlerterm unkorreliert mit X_i , kann der Erwartungswert über den bereits beschriebenen Ausdruck dargestellt werden. Ist $\gamma_{i1} = 0$, so liegt kein Selektionsproblem vor. Es muss jedoch davon ausgegangen werden, dass $\gamma_{i1} \neq 0$, da zu erwarten ist, dass die Differenz zwischen Arbeitseinkommen und Reservationslohn für Nichterwerbstätige kleiner ist als null. Es gilt:

$$\begin{aligned} E[y_{i1} | X_i, y_{i2}] &= \beta_1 X_{i1} + \gamma_{i1} E[v_{i2} | X_i, y_{i2}] \\ &= \beta_1 X_{i1} + \gamma_{i1} \Theta(X_i, y_{i2}). \end{aligned} \tag{5.10}$$

Für die am Arbeitsmarkt teilnehmenden Beobachtungen gilt $y_{i2} = 1$. Dabei kann Gleichung (5.9) einbezogen werden. Daraus ergibt sich folgende Bedingung

$$\Theta(X_i, 1) = E[v_{i2} | v_{i2} \geq -\delta_2 X_i] = \lambda_i(\delta_2 X_i).$$

Das nach John Mills¹⁷³ benannte inverse Mills Ratio kann zur weiteren Lösung verwendet werden. Dabei gilt $\lambda \equiv \frac{\phi(\delta_2 X_i)}{\Phi(\delta_2 X_i)}$ und beschreibt demnach das Verhältnis zwischen der Dichte- und der Verteilungsfunktion der spezifizierenden Merkmale für die Erwerbspersonen. Gleichung (5.10) kann umgeschrieben werden zu

¹⁷³Vgl. Heckman (1976).

$$E[y_{i1} | X_i, y_2 = 1] = \beta_1 X_{i1} + \gamma_1 \lambda_i(\delta_2 X_i). \quad (5.11)$$

Anhand der Schätzgleichung wird deutlich, dass die Selbstselektion ein Problem durch fehlende Variablen¹⁷⁴ ist. Dies hat inkonsistente Koeffizienten zur Folge. Daraus wird deutlich, dass für β_1 und γ_1 konsistente Schätzer zu berechnen sind, wenn die fehlende Variable in das Modell implementiert wird. Die Schätzung von δ_2 ist jedoch aufgrund der Unbeobachtbarkeit des Reservationslohns nicht möglich. Um dem zu begegnen schlägt Heckman (1979) folgendes Verfahren vor:

1. Zunächst gilt es, das *inverse Mills Ratio* durch eine Probit-Schätzung zu ermitteln. Dabei werden alle Erwerbspersonen einbezogen und geschätzt wird:

$$P(y_{i2} | X_i) = \Phi(\delta_2 X_i).$$

Es ergibt sich daraus $\hat{\lambda}_i \equiv \lambda_i(\hat{\delta}_2 X_i)$. Demnach wird ein Schätzwert für die unterschiedlichen Eigenschaften zwischen den Personengruppen ermittelt.

2. Im zweiten Schritt kann mit Hilfe einer GLS-Regression¹⁷⁵ die Lohngleichung für die Erwerbstätigen geschätzt werden, wobei der geschätzte Abweichungsvektor mit einbezogen wird.

Ob tatsächlich ein Selektionsproblem vorliegt, kann durch einen einfachen t-Test auf die Signifikanz von γ geprüft werden.

Erweiterter Datensatz (Erwerbspersonen) und Beschreibung der neuen Variablen. In diesem Abschnitt werden die Darstellung der Erweiterung des Datensatzes und die Beschreibung neuer Variablen diskutiert. Dabei werden deskriptive Statistiken für den erweiterten Datensatz gezeigt. Die zusätzliche Variablenbeschreibung ergibt sich dadurch, dass beim Heckman-Verfahren zunächst eine Regression auf die Arbeitsmarktbeteiligung geschätzt wird, um den notwendigen Korrekturfaktor der Lohngleichung zu ermitteln.

¹⁷⁴Auch bezeichnet als *omitted variable bias*. Vgl. Abschnitt 5.2.

¹⁷⁵Sie wird nötig um, der einzuhaltenden Heteroskedastizität gerecht zu werden. Die einfache Korrektur mit den Standardfehlern nach White ist an dieser Stelle nicht ausreichend, da die Variabilität durch die vorgelagerte Schätzung von $\hat{\delta}_2$ doppelt Einzug hält.

Erweiterung des Datensatzes. In diesem Datensatz werden neben den 4.693 Erwerbstätigen und den 119 ALG-I-Empfängern auch 436 Empfänger von ALG II berücksichtigt. Insgesamt sind dabei 5.248 Beobachtungen im Datensatz, mit 555 Erwerbslosen. Dabei ist es unbedeutend, ob die Erwerbslosen ALG I oder ALG II erhalten. Die Restriktion auf Personen im Alter zwischen 30 und 60 Jahren bleibt bestehen. Insgesamt sind in diesem Datensatz knapp 11 % der Personen nicht erwerbstätig. In Tabelle 5.28 werden ausgewählte deskriptive Statistiken für den erweiterten Datensatz dargestellt sowie diese nach den jeweiligen Erwerbsgruppen separiert.

Merkmal	Erweiterter Datensatz	Erwerbstätige	Erwerbslos
weiblich	0,48	0,47	0,52
30–40 Jahre	0,30	0,30	0,28
41–50 Jahre	0,39	0,40	0,32
50–60 Jahre	0,31	0,30	0,40
n	5.248	4.693	555

Tabelle 5.28: Deskriptive Statistiken des um die Erwerbslosen erweiterten Datensatzes gesamt und separat; Quelle: DIW, eigene Berechnungen.

Die Gruppe der Erwerbstätigen entspricht genau der bisher untersuchten Gruppe und ist daher nur der Vollständigkeit halber aufgeführt. Bemerkenswert ist, dass der Anteil der Frauen bei den Erwerbslosen signifikant¹⁷⁶ höher ist als bei den Erwerbstätigen. Auch die Altersverteilung verändert sich. So ist die am meisten vertretene Kohorte bei den Erwerbslosen die der Ältesten.

Bildungsdauer	Erwerbsstatus	Mittelwert	Levene-Test	Welch-Test
	erwerbstätig	12,76		
erwerbslos	11,24		0,000	0,000

Tabelle 5.29: P-Werte aus dem Vergleich der Bildungsdauern der Erwerbstätigen und Erwerbslosen; Quelle: DIW, eigene Berechnungen.

In Tabelle 5.29 werden die mittleren Bildungsdauern für beide Gruppen separat gezeigt und über den Welch-Test auf Gleichheit geprüft. Für die Gruppe der Erwerbstätigen ist feststellbar, dass sie mit 12,8 Jahren (2,75) eine signifikant höhere Bildungsdauer auf-

¹⁷⁶Der Levene-Test ergibt dabei, dass eine Ablehnung der Gleichheit der Varianzen nicht möglich ist. Der daraufhin durchgeführte t-Test ergibt einen P-Wert von 0,021. Demnach ist die Gleichheitshypothese mit einer Irrtumswahrscheinlichkeit von 5 % abzulehnen.

weisen als die Gruppe der Erwerbslosen. Sie haben im Mittel lediglich 11,2 Jahre (2,08) allgemeines Humankapital generiert.

Repräsentativitätsprüfung für den erweiterten Datensatz. Wie bereits in Abschnitt 4.1.3 wird als Vergleichsdatsatz zur Repräsentativitätsprüfung der Mikrozensus vom statistischen Bundesamt herangezogen. Die Prüfungsmöglichkeit ist beschränkt, da vom statistischen Bundesamt insgesamt nur wenige Daten für Erwerbslose in dem speziellen Alterssegment vorliegen.

Der Vergleich der allgemeinen Erwerbslosenquoten zeigt einen tendenziell höheren Anteil im verwendeten SOEP-Datensatz. Laut statistischem Bundesamt sind im Jahr 2008 von den Erwerbspersonen¹⁷⁷ im Alter zwischen 30 und 59 Jahren 8 % erwerbslos. Aus der SOEP-Stichprobe geht eine Erwerbslosenquote von ca. 10,5 % hervor. Diese Differenz lässt sich unter anderem durch den notwendigen Ausschluss von Beobachtungen mit fehlenden Antworten begründen. So kann aus der Gruppe der Erwerbstätigen im SOEP-Datensatz bei beispielsweise fehlender Einkommensangabe die Beobachtung nicht für die Analyse verwendet werden. Diese Restriktion ist für die Gruppe der Erwerbslosen nicht in gleichem Maß gegeben. Weiterhin ist die Anzahl der Erwerbslosen mit 555 in Relation zu den 4.693 Erwerbstätigen recht gering. Demnach kann es schon durch kleinere Änderungen zu deutlichen Verschiebungen in den relativen Häufigkeiten kommen.

Der geschlechtsspezifische Vergleich der Erwerbslosenquote zeigt in beiden Stichproben eine geringere Arbeitslosenquote für Männer, wenn auch der Unterschied im Datensatz des statistischen Bundesamtes deutlich geringer ist. So wird hier eine Arbeitslosenquote für Männer in Höhe von 7,9 %, im SOEP-Datensatz von 9,6 % gemessen. Beide Ausprägungen sind im Vergleich zum jeweiligen gesamten Datensatz geringer. Die Tendenz einer geringeren Erwerbslosenquote bei Männern kann demnach in beiden Fällen bestätigt werden.

Vergleicht man nun das Durchschnittsalter der 30–60-Jährigen aus dem SOEP-Datensatz mit den 30–59-Jährigen aus dem Mikrozensus, so ist aufgrund der beiden unterschiedlichen Intervalllängen zu erwarten, dass aus dem SOEP-Datensatz ein generell höheres mittleres Alter zu berechnen ist. Die Erwerbspersonen in diesem Datensatz sind im Mittel 45,5 Jahre alt, die Befragten des Mikrozensus sind durchschnittlich 44,4 Jahre alt. Dieses Ergebnis ist aufgrund der gesetzten Intervalle nicht überraschend. Zu vermuten ist jedoch

¹⁷⁷Hierbei wurden – zur besseren Vergleichbarkeit mit dem SOEP – die Selbstständigen nicht mit zu den Erwerbspersonen gezählt.

weiterhin, dass bei ausschließlicher Berücksichtigung der Erwerbslosen ein tendenziell höheres Alter zu berechnen ist. Dabei wird unterstellt, dass das Arbeitslosigkeitsrisiko mit dem Alter steigt. Dieses kann für beide Stichproben bestätigt werden. Im SOEP-Datensatz sind die Erwerbslosen im Mittel 46,8 Jahre alt und damit deutlich älter, als bei Betrachtung des gesamten Datensatzes. Gleiche Tendenz zeigt sich auch beim Mikrozensus. Hierbei sind die Erwerbslosen allein 44,8 Jahre alt.

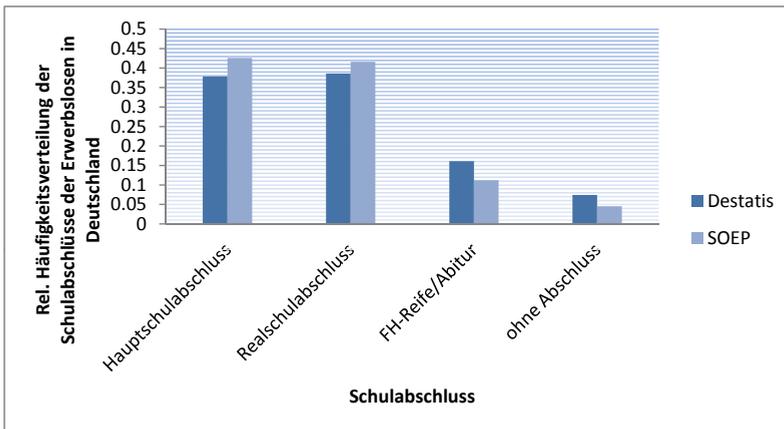


Abbildung 5.7: Relative Häufigkeitsverteilung der Schulabschlüsse der Erwerbslosen in Deutschland im Vergleich; Quelle: Destatis, DIW, eigene Berechnungen.

Der Vergleich der schulischen Bildungsabschlüsse für die Erwerbstätigen ist bereits in Abschnitt 4.1.3 gezeigt. Daher wird zunächst lediglich die relative Häufigkeitsverteilung der schulischen Abschlüsse von den Erwerbslosen verglichen. In Abbildung 5.7 werden hierfür die Anteile gegenübergestellt.

Tendenziell ist im Vergleich zu den Erwerbstätigen zu erwarten, dass die Anteile sich zugunsten der Geringqualifizierten verschieben. Im Vergleich zur Abbildung 4.1 in Abschnitt 4.1.3 ist dies deutlich erkennbar. So ist der Anteil der Personen ohne Abschluss an allen Erwerbslosen deutlich größer als es bei der Betrachtung der Erwerbstätigen der Fall war. Der Anteil der Erwerbslosen mit Fachhochschulreife bzw. Abitur ist deutlich geringer, un-

abhängig von der betrachteten Stichprobe. Im Gegensatz dazu, hat der Anteil der Personen mit Hauptschulabschluss deutlich zugenommen. Zwischen den beiden Vergleichsstichproben gibt es beim Vergleich der relativen Häufigkeiten in Bezug auf das Abitur bzw. die Fachhochschulreife Abweichungen. Es ergibt sich, dass der Anteil der Erwerbslosen mit Abitur im Datensatz des Mikrozensus mit ca. 16 % größer ist als der Anteil, der sich aus dem SOEP ergibt. Hier haben lediglich 11 % der Erwerbslosen diesen Abschluss. Alle anderen Abweichungen sind geringer. Die Tendenz, dass Erwerbslose einen im Mittel geringeren schulischen Bildungsabschluss haben als Erwerbstätige, lässt sich aber über beide Stichproben zeigen.

Fraglich ist nun, ob sich durch den höheren Anteil der Erwerbslosen im SOEP-Datensatz Verzerrungen ergeben, die gegen die Repräsentativität des Datensatzes sprechen. In Abbildung 5.8 wird daher die relative Häufigkeitsverteilung der allgemeinen Schulabschlüsse für Erwerbspersonen insgesamt verglichen. Dabei ist ersichtlich, dass die unterschiedlichen Erwerbslosenquoten eher zu einer Angleichung beider Stichproben in Bezug auf das betrachtete Merkmal führen. Daher ist für dieses Merkmal nicht davon auszugehen, dass der höhere Anteil der Erwerbslosen tendenziell zu einer gravierenden Verzerrung führt.

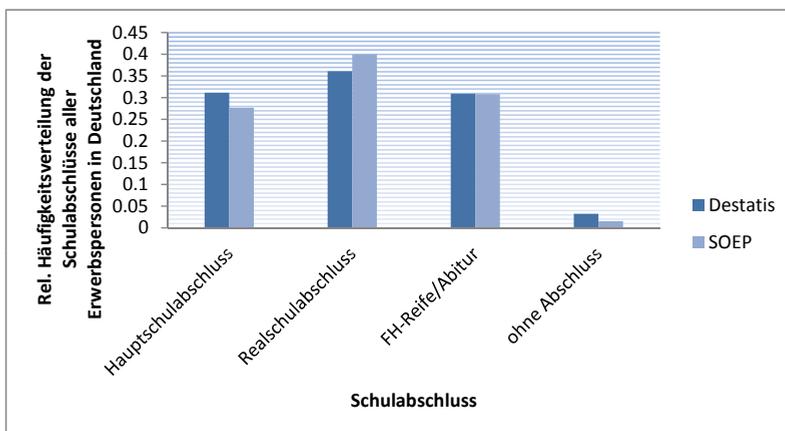


Abbildung 5.8: Relative Häufigkeitsverteilung der Schulabschlüsse der Erwerbspersonen in Deutschland im Vergleich; Quelle: Destatis, DIW, eigene Berechnungen.

Zusammenfassend ergeben sich aus der Analyse teilweise Unterschiede in den Ausprägungen der betrachteten Merkmale. Demzufolge muss für den verwendeten SOEP-Datensatz tendenziell davon ausgegangen werden, dass Erwerbslose mit einem höheren Anteil vertreten sind. Allerdings sind die vermuteten inhaltlichen Zusammenhänge der betrachteten Merkmale für beide Stichproben in gleicher Weise nachweisbar. Es gibt daher keinen Grund anzunehmen, dass dadurch die Ergebnisse erheblich verzerrend beeinflusst werden. Näherungsweise ist demnach auch hier anzunehmen – unter der Annahme, dass der Mikrozensus die Grundgesamtheit tatsächlich widerspiegelt – dass die SOEP-Daten für alle Erwerbspersonen mit Blick auf die untersuchten Merkmale ein näherungsweise repräsentatives Abbild der Wirklichkeit liefern.

Beschreibung der zusätzlichen Variablen. Zur Erklärung der Arbeitsmarktteiligung werden zu den in Abschnitt 4.2 bereits beschriebenen Variablen noch weitere hinzugefügt. Dabei handelt es sich um die Dauer der bisherigen Arbeitslosigkeit sowie die Anzahl der im Haushalt lebenden Kinder.

Dauer der Arbeitslosigkeit. Der Einfluss der Dauer der bisherigen Arbeitslosigkeit ist ein Indikator für das Arbeitslosigkeitsrisiko einer Erwerbsperson. Es ist davon auszugehen, dass bei einer Person, die bereits eine lange Dauer der Arbeitslosigkeit zu überbrücken hatte, die Wahrscheinlichkeit der Arbeitsmarktteiligung deutlich sinkt. Insgesamt liegt sie im Mittel bei 1,14 Jahren (2,51). Teilt man den Datensatz in Erwerbs- und Nichterwerbstätige, ergibt sich für Arbeitslose eine mittlere Arbeitslosigkeitsdauer von 5,59 Jahren (4,24) und für die Erwerbstätigen von 0,61 Jahren (1,51). Wie in Tabelle 5.30 ersichtlich, ist dieser deutliche Unterschied auch signifikant über den Welch-Test nachweisbar.

	Gruppierung	Mittelwert	Levene-Test	Welch-Test
Dauer der bisherigen Arbeitslosigkeit	erwerbstätig	0,61	0,000	0,000
	erwerbslos	5,59		

Tabelle 5.30: P-Werte aus dem Vergleich der Dauer der bisherigen Arbeitslosigkeit der Erwerbstätigen und Erwerbslosen; Quelle: DIW, eigene Berechnungen.

Anzahl der Kinder im Haushalt. Es kann davon ausgegangen werden, dass die Anzahl der Kinder im Haushalt einen Einfluss auf den Reservationslohn und damit auf die

Entscheidung über die Beteiligung am Arbeitsmarkt hat. Im Mittel leben in einem Haushalt 0,63 Kinder (0,01). Beim Vergleich der Ergebnisse für Erwerbstätige und Erwerbslose separat ergeben sich kaum Unterschiede im Mittelwert. So hat ein Arbeitsloser im Mittel 0,63 Kinder (1,01) und ein Erwerbstätiger im Mittel 0,62 Kinder (0,88).

Anzahl der Kinder im Haushalt	Gruppierung	Mittelwert	Levene-Test	Welch-Test
	erwerbstätig	0,62		
	erwerbslos	0,63		

Tabelle 5.31: P-Werte aus dem Vergleich der Anzahl der Kinder im Haushalt der Erwerbstätigen und Erwerbslosen; Quelle: DIW, eigene Berechnungen.

Wie die Zahlen in Tabelle 5.31 vermuten lassen, kann ein signifikanter Unterschied für dieses Merkmal zwischen Erwerbstätigen und Erwerbslosen nicht nachgewiesen werden.

Modellbeschreibung. Für die Heckman-Methode sind zwei Regressionsschritte erforderlich. Zunächst wird eine Probit-Regression durchgeführt. Für sie¹⁷⁸ wird das folgende Regressionsmodell unterstellt

$$P(y_{i2} = 1 | X_i) = \delta_0 + \delta_1 ALD_i + \delta_2 ZKH_i + \delta_3 D_{Nicht\ Deutschi} + \delta_4 S_i + \delta_5 weibl_i + \delta_6 verheii + v_i.$$

Dabei werden die Koeffizienten mit δ bezeichnet, die abhängige Variable ist die Arbeitsmarktbeteiligung. Ist eine Person erwerbstätig, erhält y_{i2} die Ausprägung 1, andernfalls 0. Für alle Erwerbspersonen sind die verwendeten Variablen beobachtbar. In Tabelle 5.32 werden sie mit vermutetem Zusammenhang zur Arbeitsmarktbeteiligung gezeigt.

¹⁷⁸Weitere Anwendungen des Verfahrens finden sich u. a. Heckman (1976); Fersterer und Winter-Ebmer (2003); Greene (2008 (6. Auflage)), S. 889 oder Wooldridge (2002), S. 565.

Kurzbezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang zur Arbeitsmarktbeteiligung
ALD_i	Dauer der Arbeitslosigkeit	metrisch	negativ
ZKH_i	Zahl der Kinder im Haushalt	metrisch	?
$D_{Nicht-Deutsch}$	Nationalität nicht Deutsch	dichotom (1: sonst, 0: Nationalität: deutsch)	negativ
S_i	Bildungsdauer	metrisch	positiv
$weibl_i$	weiblich	dichotom (1: weibl., 0: männl.)	negativ
$verhei_i$	Familienstand	dichotom (1: verh., zusammenlebend, 0: sonst)	positiv

Tabelle 5.32: Variablenbezeichnung mit vermutetem Zusammenhang zur Arbeitsmarktteilnahme.

Es ist davon auszugehen, dass sich die Dauer der Arbeitslosigkeit negativ¹⁷⁹ auf die Teilnahme am Arbeitsmarkt auswirkt. Die Richtung des Effekts der Anzahl der Kinder im Haushalt auf die Arbeitsmarktbeteiligung ist nicht eindeutig identifizierbar. Einerseits ist die Notwendigkeit der Generierung von Einkommen mit Kindern höher. Zum Teil wird dieses ausgeglichen über staatliche Transferzahlungen. Andererseits wächst mit der Kinderzahl die Unflexibilität der Erwerbsperson und die Opportunitätskosten der Arbeit steigen. Zu vermuten ist, dass der negative¹⁸⁰ Effekt überwiegt. Weiterhin kann der Einfluss einer nichtdeutschen Staatsbürgerschaft als negativ¹⁸¹ wirkender Faktor auf die Arbeitsmarktbeteiligung gewertet werden. Zudem werden die individuelle Bildungsdauer, das Geschlecht und der Familienstand mit einbezogen. Aus dieser Probit-Regression wird der Schätzer für den Korrekturfaktor $\hat{\lambda}$ berechnet.

In der Lohngleichung wird dieser berechnete Term mit einbezogen. Die Regressionsfunktion, aufbauend auf Modell 4 in Abschnitt 5.1.1, lautet dann

¹⁷⁹Vgl. auch Van den Berg (1990).

¹⁸⁰Vgl. Heckman (1979).

¹⁸¹Vgl. Miller (1966).

$$\begin{aligned} \ln y_{i1} &= \beta_0 + \beta_1 S_i + \beta_2 Erf_i + \beta_3 Erf_i^2 + \beta_4 weibl_i + \beta_5 verhei_i \\ &+ \beta_6 IATW-verh_i + \beta_7 D_{gro\beta}_i + \beta_8 D_{l\ddot{a}ndl}_i + \beta_9 D_{pendler}_i + \gamma \hat{\lambda}_i + \varepsilon_i, \end{aligned}$$

wobei die Selektionskorrektur im Term $\gamma \hat{\lambda}_i$ gezeigt wird. Da bei dieser Schätzung lediglich Erwerbstätige berücksichtigt sind, wird als abhängige Variable wieder der Bruttostundenlohn gewählt. Ob ein Selektionsproblem grundsätzlich vorliegt, zeigt die Signifikanz des Koeffizienten γ . Aus der theoretischen Überlegung in Abschnitt 5.4.1 wird die Entscheidung über die Arbeitsmarktbeteiligung als Kalkül zwischen potentielltem Arbeitseinkommen und Reservationslohn verstanden. Dass die Ungleichung 5.4 erfüllt ist, ist für Personen mit hohem potentiellen Arbeitsmarkteinkommen deutlich wahrscheinlicher. Somit wird angenommen, dass Personen, die ein tendenziell geringeres Einkommen erzielen würden, eher nicht am Arbeitsmarkt beteiligt sind. Demnach ist davon auszugehen, dass der Einfluss des Korrekturfaktors auf das Einkommen negativ ist. Alle weiteren Variablen entsprechen den in Abschnitt 5.1.1 beschriebenen.

Gesamtdeutsche Betrachtung. Für die gesamtdeutsche Regression – die Ergebnisse sind dargestellt in Tabelle 5.33 – ergibt sich im Mittel eine Verzinsung von 7,8 %¹⁸² pro zusätzlichem Bildungsjahr.

Die weiteren Koeffizienten zeigen den vermuteten Zusammenhang und entsprechen in der Höhe den Erwartungen. Der Einfluss des Korrekturterms ist, wie vermutet, negativ. Es ist demnach davon auszugehen, dass die nicht berücksichtigten Erwerbslosen tendenziell ein geringeres Einkommen erzielen würden. Die Signifikanz des Koeffizienten zeigt, dass das vermutete Selektionsproblem tatsächlich vorliegt und die mit der OLS-Schätzung berechneten Koeffizienten entsprechend verzerrt sind.

Die vorgelagerte Probit-Regression ergibt, dass mit längerer Dauer der Arbeitslosigkeit und steigender Kinderzahl die Wahrscheinlichkeit für die Arbeitsmarktbeteiligung sinkt. Im Gegenzug wird sie erhöht durch eine längere Bildungsdauer. Nicht signifikant sind die Effekte ausgehend von der nicht deutschen Staatsangehörigkeit und dem weiblichen Geschlecht, obwohl beide Koeffizienten einen negativen Koeffizienten aufweisen. Dem-

¹⁸²Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs wie folgt berechnet:
 $prozentuale\ Bildungsrendite = (e^{\beta_1} - 1) \cdot 100.$

nach können die vermuteten Zusammenhänge auf die Arbeitsmarktteiligung zumindest tendenziell bestätigt werden.

<i>Arbeitsmarktteiligung</i>		<i>Lohnleichung</i>	
Variable		Variable	
Dauer der bisherigen Arbeitslosigkeit	-0,282*** (0,010)	Bildungsjahre	0,074*** (0,002)
Staatsangehörigkeit nicht deutsch	-0,108 (0,107)	Erfahrung	0,028*** (0,004)
Anzahl der Kinder im HH	-0,082** (0,032)	Erfahrung ²	-0,000*** (0,000)
Bildungsjahre	0,064*** (0,013)	weiblich	-0,113*** (0,021)
weiblich	-0,043 (0,058)	verheiratet	0,100*** (0,018)
verheiratet	0,300*** (0,061)	IAT weiblich verheiratet	-0,151*** (0,024)
Konstante	0,951*** (0,169)	großes Unternehmen	0,211*** (0,011)
		Pendler	0,055*** (0,012)
		ländliches Gebiet	-0,050*** (0,012)
		Konstante	1,338*** (0,058)
		λ	-0,360*** (0,025)
Wald-Test n (n Arbeitslos)		0,000 5.248 (555)	

Tabelle 5.33: Ergebnistabelle der Heckman-Regression; Quelle: DIW, eigene Berechnungen; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %.

5.4.4 Kritik

Die Einbeziehung der Arbeitslosen ist aus technischen und definitorischen Gründen bei Verwendung des 3SLS-Verfahrens nur in geringem Umfang möglich. Tatsächlich zusätzlich berücksichtigt werden zunächst lediglich die Empfänger von ALG I. Die Basis, aus der sich die Höhe des ALG I bemisst, ist das zuletzt erhaltene Einkommen. Die Bestimmungsgrößen sind ähnlich wie beim Arbeitseinkommen der aktuell Erwerbstätigen, so dass in diesem Fall die Lohnleichung sinnvoll schätzbar ist. Allerdings ist der Anteil der

ALG-I-Empfänger in der erweiterten Stichprobe sehr gering, und daher ist nur eine geringe Einflussnahme auf die mittlere Bildungsrendite zu erwarten und auch zu bestätigen.

Bei der Regression der Lohngleichung muss die abhängige Variable als Nettogröße aufgenommen werden, da der ausbezahlte ALG-I-Betrag nur als solche zu verstehen ist. Die Nettobeträge können jedoch je nach familiärer Situation bei gleichem Bruttoeinkommen verschieden sein. Auch die Höhe des Auszahlungsbetrags von 60 % bzw. 67 % des vormaligen Gehalts ist abhängig von soziodemografischen¹⁸³ Faktoren. Außerdem kann keine Umrechnung in den Stundenlohn erfolgen, da die vormalige vereinbarte Arbeitszeit der ALG-I-Empfänger nicht bekannt ist. Demnach sind hierbei die Monatsbeträge aufgenommen und die Unterscheidung in Voll- und Teilzeiterwerbstätige wird vernachlässigt. Aufgrund dieser Einschränkungen ist die Vergleichbarkeit der Regressionsergebnisse mit den Ergebnissen aus dem regulären Datensatz nicht vollumfänglich gegeben. Hierfür wird eine zusätzliche Regression mit dem ursprünglichen Datensatz und der angepassten Modellspezifikation durchgeführt. Es ist festzustellen, dass sich über die Erweiterung des Datensatzes um die ALG-I-Empfänger kaum Änderungen in der Höhe der Bildungsrenditen ergeben. Demnach ist anzunehmen, dass die Veränderungen im Datensatz und der Variablenauswahl die Ergebnisse nicht wesentlich beeinflussen. Zumal der geringe zusätzliche Anteil der ALG-I-Empfänger kaum einen Einfluss auf die mittlere Rendite hat.

Da die Hinzunahme der ALG-I-Empfänger keinen wesentlichen Erkenntniszuwachs liefert, wird eine Erweiterung des Datensatzes auf alle Erwerbslosen angestrebt. Das hierfür verwendete Heckman-Verfahren berechnet im ersten Schritt einen Korrekturterm, da davon auszugehen ist, dass sich am Arbeitsmarkt Beteiligte und nicht Beteiligte unterscheiden. Die Einflüsse auf die Arbeitsmarktbeteiligung sind vielfältig. Einige Größen wurden quantifiziert und in die Probit-Regression eingebaut. Es gibt jedoch qualitative Größen, die nur schwer fassbar sind und demnach nicht als Erklärungsvariable verwendet werden können. Doch trotz zum Teil unberücksichtigter Einflussgrößen, zeigt der Korrekturfaktor das erwartete Vorzeichen. Im Gegensatz zur vorherigen Analyse erfolgt beim Heckman-Verfahren keine Korrektur innerhalb der Gruppe der Erwerbstätigen bezüglich der verschiedenen Fähigkeiten und Motivationen. Es kann jedoch angenommen werden, dass die Selbstselektion tendenziell in ähnlicher Weise wirkt und durch die Aufnahme des Korrekturfaktors auch hierfür näherungsweise kontrolliert wird.

¹⁸³Siehe gesetzliche Grundlage, Abschnitt 5.4.2.

5.4.5 Zusammenfassung

Die Berücksichtigung der Erwerbslosen zielt auf die Ausweitung der Aussagekraft der Regressionsergebnisse. Im ersten Schritt werden mit dem 3SLS-Verfahren die ALG-I-Empfänger zusätzlich im Datensatz berücksichtigt. Aufgrund gleicher Einflussgrößen des Einkommens ist die Aufnahme dieser Personengruppe sinnvoll. Aus diesen Regressionen ergibt sich eine gesamtdeutsche mittlere Bildungsrendite für Erwerbstätige und ALG-I-Empfänger in Höhe von 10,2 % und zeigt damit keine Veränderung zur vorhergehenden gesamtdeutschen Analyse auf Basis der Erwerbstätigen. Die zusätzliche Berücksichtigung der ALG-II-Empfänger über das Heckman-Verfahren ergibt in der Höhe geringere Bildungsrenditen. Es kann mit diesem Verfahren eine mittlere jährliche Verzinsung von Humankapital in Höhe von 7,8 % geschätzt werden.

6 Regionale Bildungsrenditen

Die bisherigen Analysen zeigen leichte Tendenzen zu unterschiedlichen Bildungsrenditen bei der Separierung des Datensatzes nach ausgewählten Merkmalen. Durch die deutsche Geschichte, insbesondere die Wiedervereinigung vor 18 Jahren¹⁸⁴, ist davon auszugehen, dass regionale Unterschiede bestehen. Daher werden in diesem Abschnitt regionale Aspekte in die Analyse mit einbezogen. Die Region eines Arbeitnehmers wird über den Wohnort zum Zeitpunkt der Befragung definiert.

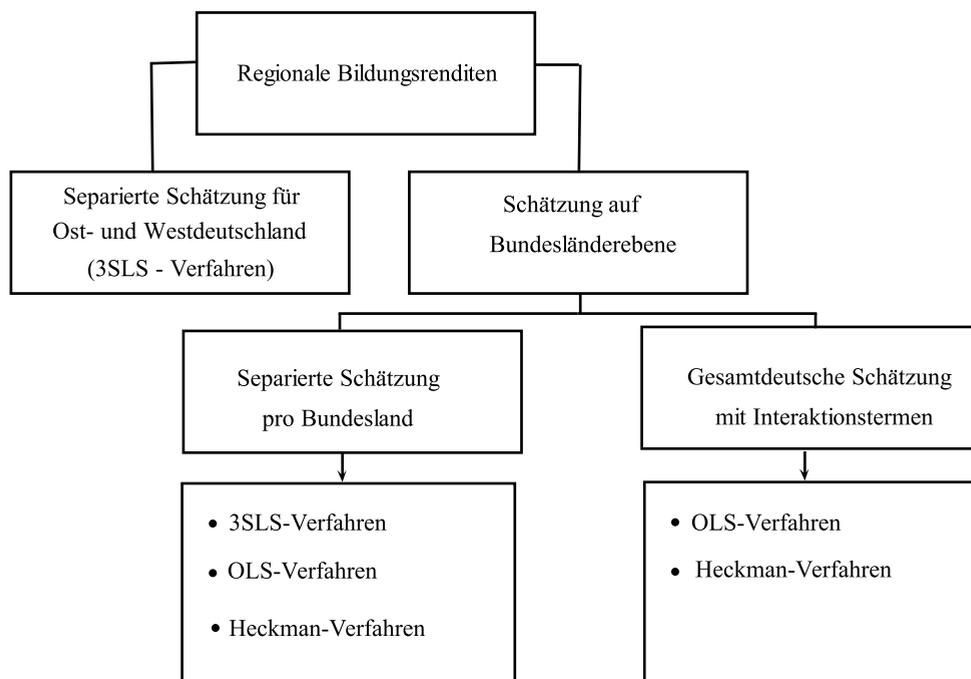


Abbildung 6.1: Übersicht Kapitel 6; eigene Darstellung.

¹⁸⁴Der Datensatz bezieht sich auf das Jahr 2008.

In Abbildung 6.1 wird der Aufbau dieses Abschnitts¹⁸⁵ im Überblick gezeigt. Nach dem Ost-West-Vergleich wird zunächst eine separierte Schätzung pro Bundesland mit dem 3SLS-Verfahren diskutiert. Die genauen Modellspezifikationen für dieses Verfahren sind bereits in den Abschnitten 5.1.1 und 5.2.1 diskutiert worden und wurden in Abschnitt 5.3 zur Regressierung der Bildungsrenditen pro Merkmal verwendet. Aus Gründen der Vergleichbarkeit der einzelnen geschätzten Bildungsrenditen über einen statistischen Test, wird im Weiteren eine Schätzung der regionalen Bildungsrenditen auf Basis einer gesamtdeutschen Stichprobe gezeigt. Diese ist jedoch lediglich mit dem OLS-Verfahren schätzbar. Daher wird als Vergleichsinstrument eine weitere separierte Schätzung der Bildungsrenditen mit dem OLS-Verfahren vorgenommen. Diese Ergebnisse können dann mit den Ergebnissen aus der nichtseparierten regionalen OLS-Schätzung verglichen werden. Zur zusätzlichen Berücksichtigung der Erwerbslosen werden die regionalen Bildungsrenditen über das Heckman-Verfahren geschätzt. Dabei findet wiederum sowohl die regionale Schätzung über den separierten Datensatz als auch über den gesamtdeutschen Datensatz mit Variation über Interaktionsterme statt.

6.1 Bildungsrenditen in Ost- und Westdeutschland

Zunächst werden die Gemeinsamkeiten und Unterschiede der Regionen durch einen Vergleich von Anteilen und Mittelwerten aufgezeigt. Nachfolgend werden dann die separierten 3SLS-Regressionen dargestellt und diskutiert. Dabei wird Ostdeutschland über den geographischen Raum definiert. Die Sonderstellung Berlins wird nicht berücksichtigt und komplett dem ostdeutschen Raum zugeordnet.

6.1.1 Deskriptive Statistiken

23 % der Arbeitnehmer dieses Datensatzes haben einen Wohnsitz in Ostdeutschland. Die Ergebnisse der relativen Häufigkeiten für ausgewählte Merkmale sind in Tabelle 6.1 dargestellt. Zugleich wird ein Signifikanztest auf Gleichheit der Häufigkeiten durchgeführt, um die Gemeinsamkeiten und Unterschiede der Personen in den jeweiligen Regionen aufzudecken.

¹⁸⁵Dieser Abschnitt basiert teilweise auf Reilich (2012).

Anteil	Ost- deutsch- land	West- deutsch- land	Levene- Test	Mittelwert- Vergleichs- Test
weiblich	0,51	0,46	0,024	0,002
öffentlicher Dienst	0,32	0,30	0,146	0,419
Vollzeitbeschäftigt	0,83	0,74	0,000	0,000
30–40 Jahre	0,30	0,30	0,752	0,875
41–50 Jahre	0,38	0,41	0,001	0,111
50–60 Jahre	0,32	0,29	0,000	0,069
n	1.082	3.611		

Tabelle 6.1: Regionaler Vergleich anhand relativer Häufigkeiten; Testergebnisse werden durch P-Werte angegeben; Quelle: DIW, eigene Berechnungen.

Wie in Abschnitt 4 werden dafür der Levene- und t- bzw. Welch-Test verwendet. Die Testergebnisse sind durch die P-Werte in Tabelle 6.1 angegeben. Auch hier gilt, wenn der P-Wert kleiner ist als $0,05$, so wird die Nullhypothese auf Gleichheit der Anteile im Vergleich beider Stichproben abgelehnt. Dabei zeigt sich, dass in Ostdeutschland 51 % der im Datensatz betrachteten Arbeitnehmer weiblich sind, im Westen sind es dagegen mit 46 % signifikant weniger. Der Vergleich der Anteile der Arbeitnehmer im öffentlichen Dienst zeigt keine signifikanten Unterschiede. Dagegen können die Anteile der Vollzeitbeschäftigten als signifikant verschieden in Ost- und Westdeutschland nachgewiesen werden. Im Osten haben 83 % der Arbeitnehmer mehr als 30 h Arbeitszeit pro Woche vertraglich vereinbart. Im Westen dagegen sind es mit lediglich 74 % deutlich weniger. Über den Welch-Test kann dabei die Gleichheitshypothese mit einer Irrtumswahrscheinlichkeit von 5 % abgelehnt werden. Die jeweiligen Altersstrukturen pro Region sind nicht signifikant verschieden. In beiden Regionen Deutschlands sind die meisten Arbeitnehmer des Datensatzes zwischen 41 und 50 Jahre alt. Die jüngste und älteste Kohorte ist jeweils mit ca. 30 % im Datensatz vertreten. Demzufolge wird deutlich, dass der Anteil der Frauen sowie der Grad der Vollbeschäftigung sich im Ost-West-Vergleich signifikant unterscheiden. Alle anderen Separierungsmerkmale zeigen keine signifikanten Unterschiede zwischen beiden Regionen.

Betrachtet man nun im nächsten Schritt die mittlere Bildungsdauer sowie den mittleren Bruttostundenlohn für die beiden Gruppen separat, so ergibt sich im Vergleich eine signifikant höhere¹⁸⁶ mittlere Bildungsdauer in Ostdeutschland.

¹⁸⁶Wie bereits in Abschnitt 4.2.2 ausgeführt, kann sich dieser Mittelwert dadurch ergeben, dass lediglich Arbeitnehmer betrachtet werden.

	Ost- deutsch- land	West- deutsch- land	Levene- Test	Mittelwert- vergleichs- Test
Durchschnittliche Bildungsdauer in Jahren	13,2 (2,562)	12,6 (2,795)	0,000	0,000
Mittlerer Bruttostundenlohn in €	14,15 (7,778)	18,71 (9,582)	0,000	0,000
Mittlere wöchentliche Arbeitszeit	37,06 (6,200)	34,60 (8,495)	0,000	0,000

Tabelle 6.2: Regionaler Vergleich der mittleren Bildungsdauer und des mittleren Bruttostundenlohns; Testergebnisse werden durch P-Werte angegeben; Quelle: DIW, eigene Berechnungen.

Im Gegensatz dazu haben jedoch Arbeitnehmer in Ostdeutschland mit 14,15 € (7,78) einen signifikant geringeren mittleren Bruttostundenlohn als Arbeitnehmer in Westdeutschland. Sie verdienen 18,71 € pro Stunde (9,58). Zur genauen Beurteilung der tatsächlichen Lohndifferenzen muss die Kaufkraftparität¹⁸⁷ berücksichtigt werden. Einen Einfluss auf die separierten Bildungsrenditen hat dies jedoch nicht, da diese pro West- und Ost-Region separat ermittelt werden. Das Ziel ist die Verzinsung der Bildung zu ermitteln, nicht das grundsätzlich verschiedene Einkommensniveau. Über den Achsenabschnitt werden zudem die unterschiedlichen Einkommensniveaus abgebildet. Schöler (2007) zeigt, dass sich die mittlere wöchentliche Arbeitszeit zwischen Ost- und Westdeutschland unterscheidet. Der Vergleich des Datensatzes bestätigt das signifikant.

6.1.2 Ergebnisdarstellung

Die Ergebnisse aus der 3SLS-Regression sind in Tabelle 6.3 dargestellt. Die Bildungsrenditen unterscheiden sich deutlich zwischen den Regionen. In Ostdeutschland wird ein zusätzliches Bildungsjahr mit 14,8 %¹⁸⁸ Einkommenssteigerung entlohnt, wohingegen Arbeitnehmer mit westdeutschem Wohnsitz lediglich 10,1 % pro zusätzlichem Bildungsjahr erhalten.

¹⁸⁷Siehe hierfür Goebel et al. (2009).

¹⁸⁸Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs wie folgt berechnet:

$$\text{prozentuale Bildungsrendite} = (e^{\beta_1} - 1) \cdot 100.$$

6.1 Bildungsrenditen in Ost- und Westdeutschland

Bildungs- gleichung	Ostdeutsch- land	Westdeutsch- schland	Lohnglei- chung	Ostdeutsch- land	Westdeutsch- schland
mittl. väterl. Bildung	1,450*** (0,135)	1,814*** (0,086)	Bildungs- jahre	0,138*** (0,009)	0,096*** (0,004)
hohe väterl. Bildung	3,748*** (0,241)	4,523*** (0,169)	Erfahrung	0,028*** (0,007)	0,032*** (0,004)
Anz. d. Bücher	1,577*** (0,135)	1,523*** (0,083)	Erfahrung ²	-0,000*** (0,000)	-0,000*** (0,000)
Anz. d. Schwestern	-0,259** (0,063)	-0,152*** (0,036)	weiblich	-0,066 (0,041)	-0,150*** (0,022)
Anz. der Brüder	-0,112* (0,061)	-0,237*** (0,035)	verheiratet	0,066* (0,038)	0,109*** (0,019)
Staatsan. nicht deutsch	-4,026*** (1,154)	-1,058*** (0,147)	IAT verh. Frau	-0,087* (0,050)	-0,143*** (0,027)
Konstante	16,279*** (1,160)	12,902*** (0,169)	großes Un- ternehmen	0,257*** (0,024)	0,169*** (0,012)
			ländliches Gebiet	-0,025 (0,024)	-0,019 (0,013)
			Pendler	0,040* (0,024)	0,061*** (0,012)
			Konstante	0,239 (0,121)	1,162*** (0,088)
korr. R ²	0,35	0,36	korr. R ²	0,35	0,37
Wald-Test	0,000	0,000	Wald-Test	0,000	0,000
n	1.082	3.611	n	1.082	3.611

Tabelle 6.3: Ergebnistabelle der 3SLS-Regression für Ost- und Westdeutschland; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

Ebenso wie in Abschnitt 5.3 können die Koeffizienten über einen statistischen Test nicht verglichen werden. Als Indiz für die tatsächliche Unterschiedlichkeit werden daher die 95 %-Konfidenzintervalle herangezogen. In Tabelle 6.4 sind diese dargestellt. Der Vergleich ergibt keinerlei Überschneidung der Intervalle. Demzufolge kann davon ausgegangen werden, dass die Unterschiede der Bildungsrenditen nicht zufällig sind. Es ist demnach anzunehmen, dass sich für die 30–60-Jährigen Arbeitnehmer in Ostdeutschland ein zusätzliches Bildungsjahr deutlich höher verzinst als in Westdeutschland. Das grundsätzlich unterschiedliche Einkommensniveau bestätigt die beiden Konstanten der Lohngleichung.

	95 %-Konfidenzintervall
Ostdeutschland	0,129–0,156
Westdeutschland	0,088–0,104

Tabelle 6.4: 95 %-Konfidenzintervalle für die Bildungsrendite in Ost- und Westdeutschland; Quelle: DIW, eigene Berechnungen.

Betrachtet man die weiteren Kontrollvariablen, so ergeben sich ähnliche Koeffizienten in beiden Regionen für den Einfluss der Berufserfahrung auf das Einkommen. Die anderen Koeffizienten der Lohnleichung sind zum Teil nicht statistisch gesichert. So wirken in Ostdeutschland weder das weibliche Geschlecht, noch der eheliche Familienstand, noch die Kombination aus beidem nachweislich auf den logarithmierten Bruttostundenlohn. Die Koeffizienten für Westdeutschland sind signifikant und entsprechen den Erwartungen. Der Einfluss der Unternehmensgröße ist dagegen für beide Unterstichproben statistisch relevant nachweisbar, wobei der Koeffizient für Ostdeutschland deutlich größer ist. Der Einfluss eines ländlichen oder urbanen Gebietes kann für beide Stichproben separiert nicht als signifikanter Einfluss für das Einkommen ermittelt werden. Anders hingegen verhält es sich beim Pendlerverhalten. Hierfür ist ein signifikant positiver Koeffizient für Westdeutschland berechenbar. In Ostdeutschland dagegen kann dies mit einer Irrtumswahrscheinlichkeit von 5 % nicht bestätigt werden.

Die Inferenzstatistik zeigt ähnliche Zahlen für beide Unterstichproben. So können 35 % der gesamten Streuung des logarithmierten Bruttostundenlohns mit dem Modell in Ostdeutschland, 37 % in Westdeutschland erklärt werden. Auch über den Wald-Test kann das Modell bestätigt werden.

Die Betrachtung der Regressionskoeffizienten der Bildungsdauern im Vergleich zwischen Ost- und Westdeutschland zeigen tendenziell ähnliche Einflüsse. Deutlich ist, dass der Einfluss der väterlichen Bildung in Ostdeutschland geringer ist. Die Anzahl der Schwestern wirkt auf die Bildungsdauer in Ostdeutschland stärker mindernd als in Westdeutschland. Der Einfluss der Zahl der Brüder ist hier lediglich auf dem 10 %-Signifikanzniveau zu bestätigen. Besonders deutlich hingegen tritt der Einfluss der nichtdeutschen Staatsbürgerschaft hervor. Zwar haben die Koeffizienten sowohl für Ost- als auch für Westdeutschland einen signifikanten Einfluss auf die Bildungsdauer, der negative Effekt ist jedoch im Osten um einiges größer. Die Inferenzstatistik zeigt einen Erklärungsgehalt der Streuung der Bildungsdauer von 35 % bzw. 36 %. Auch der Wald-Test zeigt ein weiteres Mal die Übertragbarkeit des Modells.

Mit Blick auf die Bildungsrendite zeigt sich eine große Unterschiedlichkeit in der Wert-schätzung der Bildung zwischen Ost- und Westdeutschland. Diesem näher auf den Grund zu gehen ist Aufgabe des nächsten Abschnitts. Hierbei werden auf Bundesländerebene weitere Heterogenitäten untersucht.

6.1.3 Zusammenfassung

Bildungsrenditen unterscheiden sich regional in Deutschland. In einer ersten Analyse werden die Ost-West-Regionen Deutschlands untersucht. Der Vergleich der Verzinsung für ein zusätzliches Bildungsjahr zeigt mit 14,8 % eine höhere Einkommenssteigerung in den Ländern der ehemaligen DDR. In Westdeutschland liegt dagegen die Bildungsrendite lediglich bei 10,1 % pro zusätzlichem Bildungsjahr. Der Vergleich der 95 %-Konfidenzintervalle zeigt eine deutliche Überschneidungsfreiheit. Es kann damit von einer tendenziell unterschiedlichen Bildungsrendite in Ost- und Westdeutschland ausgegangen werden.

6.2 Regionale Bildungsrenditen

Die Ergebnisse des vorangegangenen Abschnitts deuten auf unterschiedliche Bildungsrenditen in Ost- und Westdeutschland hin. Dieser Abschnitt weitert die Analyse auf Bundesländerebene. Hierfür werden verschiedene Methoden zur Schätzung regionale Renditen diskutiert. Dabei wird zunächst der Datensatz in verschiedene regionale Teildatensätze separiert, um eine jeweilige Bildungsrendite zu ermitteln. Im nächsten Schritt werden die regional verschiedenen Bildungsrenditen über die Berücksichtigung von Interaktionstermen in einer nicht separierten Stichprobe geschätzt. Ein Vergleich, der die Ergebnisse aus den verschiedenen Verfahren gegenüberstellt folgt, ebenso wie die kritische Auseinandersetzung. Dieser Abschnitt schließt mit einem zusammenfassenden Fazit.

6.2.1 Regionale Bildungsrenditen auf Basis separierter Stichproben

In diesem Teil der Arbeit werden regionale Unterschiede auf Bundesländerebene aufgezeigt. Zunächst erfolgt dafür eine Präsentation der deskriptiven Statistiken. Im zweiten Schritt wird das 3SLS-Verfahren für die jeweiligen Regionen mit den Modellspezifikationen aus den Abschnitten 5.1.1 und 5.2.1 durchgeführt. Soweit es möglich ist, wird ein

Vergleich der Renditen zwischen den Bundesländern angestrebt. Die Bundesländer werden im Folgenden abgekürzt nach den ISO-3166-2-Codes¹⁸⁹ für Deutschland.

Deskriptive Analyse des regionalen Datensatzes. Bei der Separierung des Datensatzes auf Bundesländerebene ergeben sich zum Teil nur kleine Stichprobenumfänge. Daher sind im Datensatz das Saarland und Rheinland-Pfalz von vornherein vom DIW zusammengefasst worden. In Tabelle 6.6 werden ausgewählte deskriptive Statistiken pro Bundesland angegeben.

Bundesland	n	weiblich	öffentl. Dienst	Vollzeit	30–40 Jahre	41–50 Jahre	51–60 Jahre
BW	588	0,41	0,24	0,78	0,33	0,39	0,28
BY	677	0,47	0,29	0,73	0,31	0,41	0,28
BE	133	0,57	0,38	0,75	0,23	0,42	0,35
BB	186	0,52	0,42	0,82	0,31	0,32	0,37
HB	41	0,51	0,29	0,73	0,32	0,42	0,27
HH	59	0,46	0,32	0,78	0,31	0,36	0,34
HE	349	0,51	0,32	0,72	0,31	0,42	0,27
MV	109	0,56	0,36	0,81	0,21	0,42	0,37
NI	412	0,46	0,31	0,73	0,31	0,41	0,28
NW	971	0,45	0,31	0,76	0,30	0,40	0,30
RP / SL	310	0,48	0,35	0,71	0,24	0,45	0,31
SN	349	0,50	0,26	0,80	0,32	0,39	0,29
ST	195	0,51	0,32	0,86	0,32	0,32	0,36
SH	128	0,49	0,32	0,73	0,27	0,37	0,37
TH	186	0,49	0,26	0,83	0,30	0,42	0,29

Tabelle 6.6: Regionaler Vergleich anhand relativer Häufigkeiten pro Bundesland; Quelle: DIW, eigene Berechnungen.

Zunächst gilt es die jeweiligen Stichprobenumfänge zu betrachten. Zur Durchführung des Verfahrens werden genügend große Stichprobenumfänge gefordert. Dabei wird „genügend“ von Backhaus et al. (2006 (11. Auflage)) mit ca. 40 Beobachtungen quantifiziert.

¹⁸⁹Herausgegeben von der Internationalen Organisation für Standardisierungen (ISO). So gilt für BW: Baden-Württemberg, BY: Bayern, BE: Berlin, BB: Brandenburg, HB: Bremen, HH: Hamburg, HE: Hessen, MV: Mecklenburg-Vorpommern, NI: Niedersachsen, NW: Nordrhein-Westfalen, RP/SL: Rheinland-Pfalz und Saarland, SN: Sachsen, ST: Sachsen Anhalt, SH: Schleswig-Holstein und TH: Thüringen. Die Abkürzungen sind zusätzlich im Symbolverzeichnis dargestellt.

Die kleinste Beobachtungszahl ist in Bremen mit 41 Beobachtungen gegeben. Erwartungsgemäß sind die meisten Personen in der Stichprobe aus Nordrhein-Westfalen mit 971 Beobachtungen. Mit diesen regionalen Stichprobenumfängen kann davon ausgegangen werden, dass auch nach Separierung die geforderte Beobachtungszahl aus den vorangegangenen Abschnitten gegeben ist.

Die Anteile pro Bundesland und jeweiligem Merkmal in Tabelle 6.6 werden aus Gründen der Übersichtlichkeit nicht jeweils gegeneinander über einen statistischen Test abgesichert. Gerade bei den Bundesländern mit kleinen Stichprobenumfängen können die relativen Häufigkeiten nur als grobes Indiz verstanden werden. Zur besseren Vergleichbarkeit sind die deskriptiven Statistiken in Abbildung 6.2 anhand von Deutschlandkarten¹⁹⁰ mit nach der Höhe der Bildungsrendite farblich abgesetzten Bundesländern veranschaulicht. Dabei ist zu beachten, dass die Intervalle je nach Ausprägungen gewählt wurden, um die regionale Heterogenität klarer herauszustellen. In Abbildung 6.2.a ist der Anteil weiblicher Arbeitnehmer pro Bundesland dargestellt. Dabei liegt die Spanne zwischen 41 % in Baden-Württemberg und 57 % in Berlin. Abbildung 6.2.b zeigt den Anteil der Arbeitnehmer im öffentlichen Dienst. Im Vergleich zu den anderen Bundesländern sind in Brandenburg mit 42 % die meisten Arbeitnehmer im öffentlichen Dienst beschäftigt. Baden-Württemberg verfügt hingegen über den geringsten Anteil der Arbeitnehmer im öffentlichen Dienst. Betrachtet man dagegen die Arbeitnehmer mit einer Vollzeitstelle, so haben 71 % der Arbeitnehmer in Rheinland-Pfalz mehr als 30 Stunden pro Woche vertraglich vereinbart und sind damit das Bundesland mit den meisten Teilzeitbeschäftigten. Die meisten Vollzeitkräfte dagegen wohnen in Sachsen-Anhalt mit einem Anteil von 86 % der gesamten Arbeitnehmer. Die Abbildungen 6.2.d-f zeigen die Altersstruktur pro Bundesland. Dabei wird der jeweilige Anteil der (a) 30–40-Jährigen, (b) 41–50-Jährigen und (c) 51–60-Jährigen gezeigt. Es zeigen sich regional relativ unterschiedliche Verteilungen. Aus der ersten Grafik wird deutlich, dass in Mecklenburg-Vorpommern mit 21 % der geringste Anteil der 30–40-Jährigen Arbeitnehmer wohnt, gefolgt von Berlin mit 23 % und Rheinland-Pfalz und dem Saarland mit 24 %. Der höchste Anteil dieser jüngsten Kohorte pro Bundesland ist in Baden-Württemberg. Die wenigsten Arbeitnehmer im mittleren Alter dieser separierten Stichprobe leben in Brandenburg und Sachsen-Anhalt mit jeweils 32 %, der höchste Anteil ist in Rheinland-Pfalz. Tendenziell ist jedoch hierbei ersichtlich, dass sich Brandenburg und Sachsen-Anhalt im Vergleich zu den anderen Bundesländern deutlich abgrenzen lassen. Die älteste Kohorte hat in Mecklenburg-Vorpommern mit 37 % relativ zu den anderen Bundesländern den höchsten Anteil der 51–60-Jährigen. Auffällig

¹⁹⁰Erzeugt wurden die Deutschlandkarten mit dem Programm R.

6 Regionale Bildungsrenditen

ist, dass im Nordosten Deutschlands der Anteil der ältesten Kohorte deutlich größer ist als in den anderen Bundesländern.

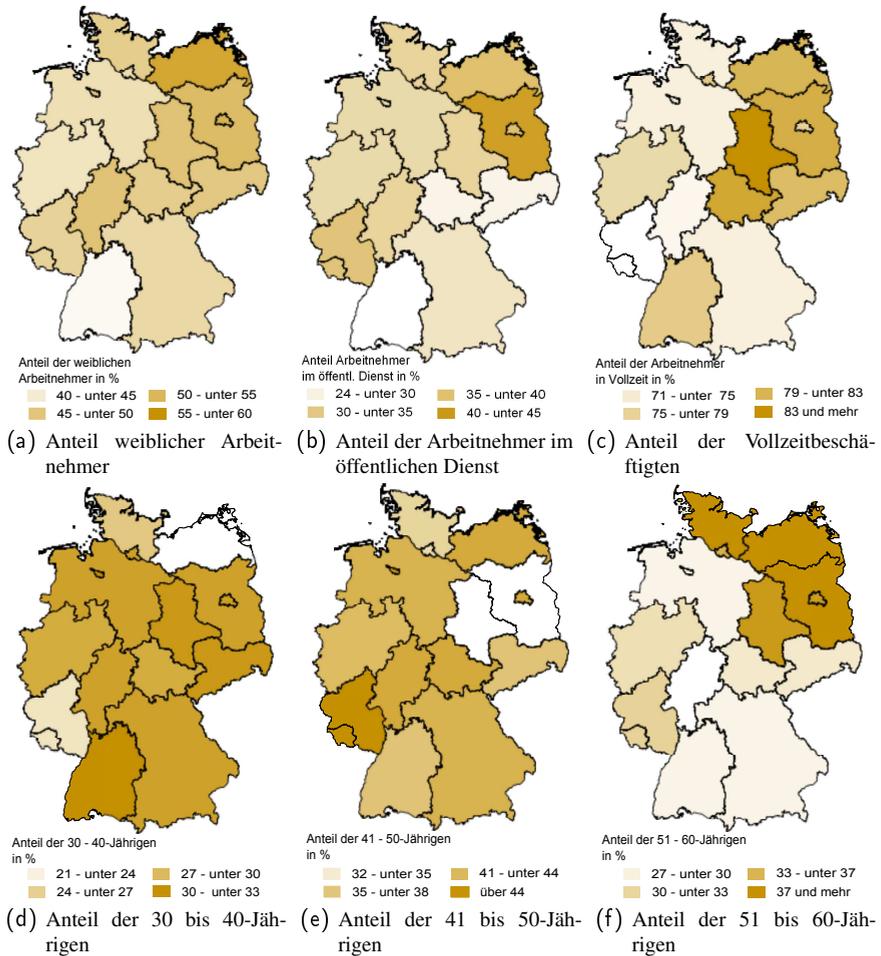


Abbildung 6.2: Vergleich relative Häufigkeiten auf Bundesländerebene; Quelle: DIW, eigene Berechnungen.

Wie in Abschnitt 6.1 werden im Weiteren auch ausgewählte Mittelwerte analysiert. Die separierten Ergebnisse sind in Tabelle 6.8 sowie in den Abbildungen 6.3a-c dargestellt. Betrachtet man zunächst die mittlere Bildungsdauer pro Bundesland, so zeigt sich eine Spanne zwischen im Mittel 12,1 Bildungsjahren (2,53) in Rheinland-Pfalz und dem Saar-

land und Hamburg mit 14,3 Bildungsjahren (3,14). Tendenziell sind die mittleren Bildungsdauern in den Stadtstaaten höher als in den Flächenländern.

Wie bereits durch den Ost-West-Vergleich dargestellt, haben die Arbeitnehmer in ostdeutschen Bundesländern tendenziell längere Bildungsdauern. Genau das gegenteilige Bild zeigt nun der Vergleich der mittleren Bruttostundenlöhne.

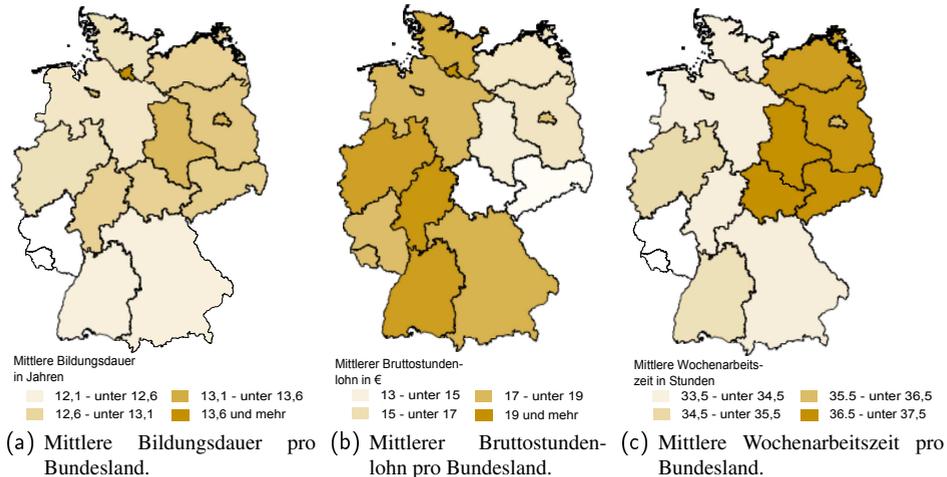


Abbildung 6.3: Mittelwertvergleich pro Bundesland; Quelle: DIW, eigene Berechnungen.

Auch dieses Ergebnis schließt sich der deskriptiven Analyse der Ost-West-Regionen an und zeigt für alle Bundesländer in Ostdeutschland geringere Bruttostundenlöhne. Einzig Berlin ist mit 17,72 € (9,73) den mittleren Bruttostundenlöhnen der Westregionen deutlich näher. Die Spannweite liegt insgesamt zwischen 13,32 € pro Stunde (6,50) in Thüringen und 20,16 € pro Stunde (6,50) in Hamburg. Dagegen ist die mittlere vereinbarte wöchentliche Arbeitszeit in den östlichen Bundesländern höher als in den Westregionen. Die geringste Ausprägung findet sich in Rheinland-Pfalz und im Saarland mit lediglich 33,9 vereinbarten Stunden (8,80).

Aus dem Vergleich der Anteile in Tabelle 6.6 und Abbildung 6.6c, wobei der Anteil der Vollzeitbeschäftigten im regionalen Vergleich am geringsten ist, entspricht dieses Ergebnis den Erwartungen. Die meiste Zeit arbeitet ein Arbeitnehmer in Thüringen und Sachsen-Anhalt mit im Mittel 37,2 Stunden pro Woche (6,01 / 6,62). Fraglich ist, ob diese Merkmale einen Einfluss auf die Wertschätzung eines zusätzlichen Bildungsjahres haben. Dass

es auch 18 Jahre nach der Wiedervereinigung Deutschlands noch deutliche Unterschiede zwischen den Regionen kann mit dieser groben Analyse bestätigt werden.

Bundesland	mittlere Bildungsdauer	mittlerer Bruttostundenlohn	mittlere Wochenarbeitszeit	Bundesland	mittlere Bildungsdauer	mittlerer Bruttostundenlohn	mittlere Wochenarbeitszeit
BW	12,4 (2,87)	19,35 (8,30)	34,8 (8,50)	NI	12,6 (2,60)	17,70 (10,09)	34,4 (8,84)
BY	12,4 (2,79)	18,11 (9,26)	34,4 (8,82)	NW	12,7 (2,78)	19,20 (10,62)	35,0 (8,03)
BE	13,9 (2,86)	17,72 (9,73)	35,8 (7,23)	RP /	12,1 (2,53)	17,34 (8,37)	33,9 (8,80)
BB	13,2 (2,51)	14,94 (7,91)	37,0 (6,08)	SN	13,1 (2,55)	13,57 (6,84)	37,1 (5,90)
HB	13,5 (3,05)	16,10 (6,62)	34,9 (8,71)	ST	13,5 (2,70)	14,47 (8,64)	37,2 (6,62)
HH	14,3 (3,14)	20,16 (11,32)	35,1 (7,33)	SH	12,7 (2,59)	18,51 (7,39)	34,4 (8,78)
HE	13,0 (2,86)	19,81 (9,63)	34,4 (8,62)	TH	13,0 (2,49)	13,32 (6,50)	37,2 (6,01)
MV	13,0 (2,42)	14,90 (10,57)	36,8 (6,86)				

Tabelle 6.8: Mittelwertvergleich auf Bundesländerebene; Quelle: DIW, eigene Berechnungen.

Ergebnisdarstellung. Die Regressionsergebnisse pro Bundesland für die Schätzung der Lohngleichung sind in den Tabellen 6.9 und 6.10 dargestellt. Bei der 3SLS-Methode ergeben sich zudem die Schätzergebnisse für die Regression auf die Bildungsdauer. Diese werden im Anhang A.4.1 in den Tabellen A.17 und A.18 gezeigt.

Mit Blick auf die Vielzahl der Koeffizienten und dem eigentlichen Forschungsgegenstand werden die berechneten Bildungsrenditen in Abbildung¹⁹¹ 6.4 zusammengefasst. Beim Vergleich der Renditen fällt auf, dass in den ostdeutschen¹⁹² Bundesländern eine deutlich höhere mittlere Verzinsung erzielt wird als in den westdeutschen. Dabei ist Mecklenburg-

¹⁹¹Das dabei gewählte Intervall ist für alle im Weiteren folgenden Deutschlandkarten identisch, um die Vergleichbarkeit der Bildungsrenditen zwischen den Karten zu gewährleisten.

¹⁹²Auch Westberlin zählt dazu.

Vorpommern mit 18,6 %¹⁹³ zusätzlichem Einkommen pro zusätzlichem Bildungsjahr Spitzenreiter. Abgesehen von Hamburg – der Koeffizient ist nur mit einer Irrtumswahrscheinlichkeit von 10 % signifikant – ist die niedrigste Bildungsrendite bei 8,2 % in Schleswig-Holstein auszumachen. Die niedrigste Rendite in den neuen Bundesländern findet sich in Berlin mit 11,2 %. Die höchste im westdeutschen Raum liegt bei 13,2 % im Stadtstaat Bremen. Der Vergleich der Renditen zueinander wird im nachfolgenden Abschnitt 6.2.1 vorgenommen.

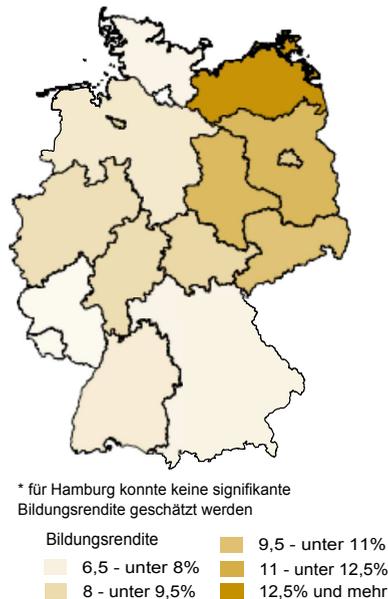


Abbildung 6.4: Bildungsrenditen pro Bundesland (3SLS-Methode); Quelle: DIW, eigene Berechnungen.

Die weiteren Koeffizienten sind aufgrund der geringen Fallzahlen häufig insignifikant. Bei gegebener Signifikanz ist jedoch stets der vermutete Zusammenhang zu bestätigen. Abgesehen von dem Koeffizienten für die Bildungsrendite sind zumeist die Koeffizienten der Dauer der Berufserfahrung und der Unternehmensgröße signifikant. Diese beiden Größen scheinen daher für die meisten Regionen einen deutlichen und wichtigen Erklärungsbei-

¹⁹³Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs wie folgt berechnet:
 $prozentuale\ Bildungsrendite = (e^{\beta_1} - 1) \cdot 100.$

6 Regionale Bildungsrenditen

Regressoren	BW	BY	BE	BB	HB	HH	HE	MV
Bildungs- jahre	0,088*** (0,008)	0,087*** (0,010)	0,106*** (0,025)	0,156*** (0,028)	0,124*** (0,021)	0,042* (0,025)	0,107*** (0,011)	0,172*** (0,026)
Erfahrung	0,033*** (0,008)	0,050*** (0,010)	0,057*** (0,022)	0,009 (0,019)	0,040 (0,025)	-0,017 (0,029)	0,023** (0,011)	0,039 (0,029)
Erfahrung ²	-0,001*** (0,000)	-0,001*** (0,000)	-0,001** (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,001 (0,001)	-0,000 (0,000)	-0,000 (0,001)
weiblich	-0,177*** (0,047)	-0,149*** (0,052)	-0,228* (0,235)	-0,027 (0,109)	-0,127 (0,142)	-0,038 (0,145)	-0,106 (0,073)	-0,107 (0,141)
verheiratet	0,124** (0,039)	0,062 (0,045)	0,208 (0,133)	0,079 (0,106)	0,004 (0,128)	0,431*** (0,134)	0,056 (0,065)	0,059 (0,117)
IAT verh.	-0,139** (0,057)	-0,120* (0,064)	0,190 (0,174)	-0,118 (0,132)	-0,031** (0,176)	-0,459** (0,192)	-0,089 (0,083)	0,016 (0,160)
Frau	0,128*** (0,026)	0,138*** (0,030)	0,350*** (0,084)	0,190 (0,065)	0,104 (0,090)	0,106 (0,100)	0,189*** (0,037)	0,116 (0,074)
großes ländliches Gebiet	-0,047* (0,027)	0,072** (0,030)	0,076 (0,153)	0,038 (0,063)	0,110 (0,168)	0,055 (0,183)	-0,035 (0,037)	0,067 (0,075)
Pendler	0,064** (0,027)	0,030 (0,030)	0,082 (0,124)	0,114 (0,068)	0,020 (0,126)	0,147 (0,140)	0,138*** (0,039)	-0,051 (0,074)
Konstante	1,420*** (0,184)	1,073*** (0,218)	0,758 (0,469)	0,124 (0,454)	0,520 (0,446)	2,156*** (0,539)	1,097*** (0,262)	-0,341 (0,565)
R ²	0,44	0,35	0,32	0,24	0,48	0,38	0,36	0,44
Wald-Test	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
n	588	677	133	186	41	59	349	109

Tabelle 6.9: Ergebnistabelle der 3SLS-Regression auf Bundesländerebene, Teil 1; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

6 Regionale Bildungsrenditen

Regressoren	NI	NW	RP / SL	SN	ST	SH	TH
Bildungsjahre	0,093*** (0,013)	0,106*** (0,009)	0,092*** (0,015)	0,137*** (0,016)	0,138*** (0,019)	0,079*** (0,017)	0,112*** (0,019)
Erfahrung	0,040*** (0,012)	0,023*** (0,008)	0,040*** (0,016)	0,028** (0,011)	0,043** (0,017)	0,043*** (0,016)	0,004 (0,017)
Erfahrung ²	-0,001** (0,000)	-0,000 (0,000)	-0,001** (0,000)	-0,000* (0,000)	-0,001** (0,000)	-0,001** (0,000)	0,000 (0,000)
weiblich	-0,204*** (0,071)	-0,178*** (0,046)	-0,186** (0,095)	-0,060 (0,069)	-0,067* (0,103)	-0,108 (0,088)	-0,042 (0,083)
verheiratet	0,192*** (0,056)	0,090** (0,038)	0,089 (0,068)	0,019 (0,059)	0,225** (0,093)	0,032 (0,085)	0,131* (0,079)
IAT verh. Frau	-0,068 (0,084)	-0,141*** (0,053)	-0,188** (0,095)	-0,074 (0,082)	-0,146 (0,122)	-0,118 (0,110)	-0,239** (0,105)
großes Untern.	0,156 (0,039)	0,185*** (0,024)	0,232*** (0,043)	0,242*** (0,041)	0,270*** (0,060)	0,144*** (0,051)	0,352*** (0,052)
ländliches Gebiet	-0,045 (0,039)	-0,023 (0,028)	-0,053 (0,042)	-0,029 (0,038)	-0,054 (0,060)	-0,028 (0,054)	-0,050 (0,052)
Pendler	0,029 (0,040)	0,053** (0,024)	0,094** (0,045)	-0,008 (0,039)	0,077 (0,060)	0,021 (0,057)	0,004 (0,054)
Konstante	1,057*** (0,260)	1,172*** (0,184)	1,129*** (0,327)	0,333 (0,294)	-0,020 (0,354)	1,265*** (0,355)	0,805** (0,315)
R ²	0,35	0,39	0,44	0,29	0,43	0,42	0,50
Wald-Test	0,000	0,000	0,000	0,000	0,000	0,000	0,000
n	412	971	310	349	195	128	186

Tabelle 6.10: Ergebnistabelle der 3SLS-Regression auf Bundesländerebene, Teil 2; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

trag des Bruttostundenlohns zu liefern. Die Ergebnisse der Inferenzstatistik zeigen eine Spanne des R^2 von 24 % in Brandenburg bis 50 % in Thüringen. Der Wald-Test kann für alle regionalen Modelle abgelehnt werden.

Vergleich zwischen den Regionen über Konfidenzintervalle. In Abschnitt 6.2.1 werden Bildungsrenditen mit der 3SLS-Methode mit separierten Stichproben pro Bundesland geschätzt. Die Ergebnisse können jedoch aufgrund der verschiedenen Stichproben nicht über einen Test miteinander verglichen werden. Um dennoch einen Vergleich der regionalen Bildungsrenditen durchzuführen, werden, wie bereits in den vorhergehenden Abschnitten, die 95 %-Konfidenzintervalle der Bildungsrenditen miteinander verglichen. In Tabelle 6.12 wird die jeweilig geschätzte Bildungsrendite und das zugehörige 95 %-Konfidenzintervall des Schätzers pro Bundesland dargestellt.

Bundesland	Bildungsrendite	Konfidenzintervall	Bundesland	Bildungsrendite	Konfidenzintervall
BW	0,088***	0,073–0,104	NI	0,093***	0,068–0,119
BY	0,087***	0,067–0,106	NW	0,106***	0,089–0,123
BE	0,106***	0,057–0,154	RP / SL	0,092***	0,063–0,121
BB	0,156***	0,101–0,211	SN	0,137***	0,105–0,169
HB	0,124***	0,083–0,165	ST	0,138***	0,100–0,176
HH	0,042*	-0,006–0,091	SH	0,079***	0,045–0,113
HE	0,107***	0,085–0,129	TH	0,113***	0,076–0,149
MV	0,172***	0,122–0,223			

Tabelle 6.12: 95 %-Konfidenzintervalle aus der 3SLS-Regression auf Bundesländerebene; Quelle: DIW, eigene Berechnungen. Signifikanzniveau: *** 1 %, ** 5 % und * 10 %.

Zur besseren Vergleichbarkeit werden die Intervalle zusätzlich in Abbildung 6.5 als Zahlenstrahl gegenübergestellt. Hierbei stellen die Intervalle oberhalb der Achse die Vertrauensbereiche aus den ostdeutschen Bundesländern dar, unterhalb aus den westdeutschen Bundesländern. Der Vergleich zeigt eine deutlich größere Spannweite der 95 %-Konfidenzintervalle für die neuen Bundesländer. Ausnahmen bilden im westdeutschen Raum die Stadtstaaten Hamburg und Bremen. Die Bildungsrendite für Hamburg ist mit einer Irrtumswahrscheinlichkeit von 5 % in diesem Modellrahmen nicht signifikant von null verschieden. Die hohen Spannweiten ergeben sich durch kleine Stichprobenumfänge, wodurch einzelne Abweichungen einen größeren Einfluss haben. Im Gegensatz dazu sind

die 95 %-Konfidenzintervalle der Bildungsrenditen für die einwohnerstarken Bundesländer erwartungsgemäß¹⁹⁴ deutlich geringer.

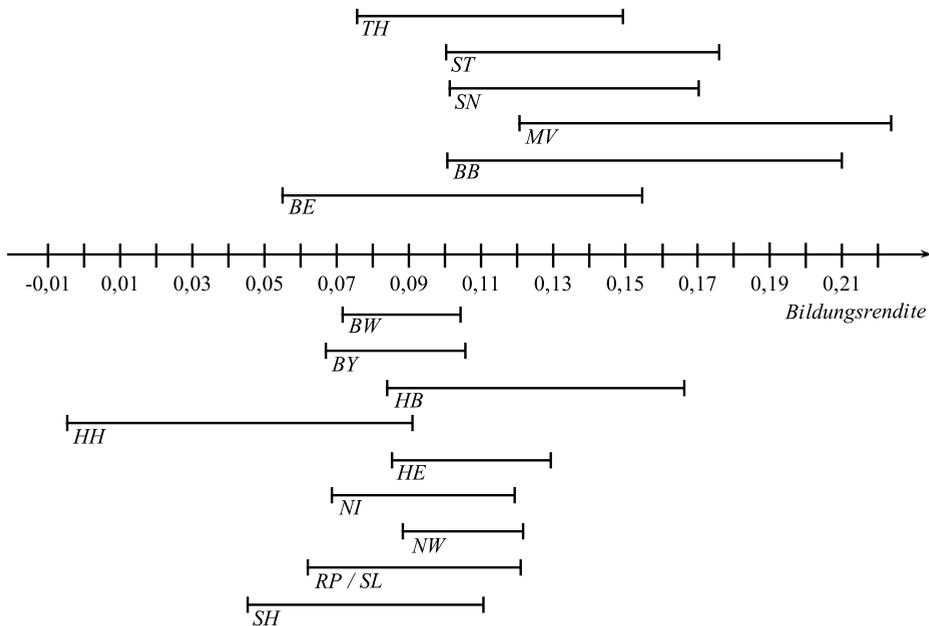


Abbildung 6.5: 95 %-Konfidenzintervalle für die Bildungsrenditen pro Bundesland (3SLS-Verfahren); Quelle: DIW, eigene Darstellung.

Auffällig ist weiterhin, dass sich die Intervalle der alten Bundesländer im Vergleich untereinander deutlich überschneiden. Auch im Ost-West-Vergleich überschneiden sich fast alle Vertrauensbereiche. In diesen Fällen kann nicht von tatsächlich verschiedenen Bildungsrenditen ausgegangen werden. Einzig der Vergleich der Vertrauensbereiche der Bildungsrendite mit Mecklenburg-Vorpommern führt zur Abgrenzbarkeit der Renditen. Im Vergleich mit den westdeutschen Bundesländern kann die Überschneidung mit dem Intervall für Mecklenburg-Vorpommern für fast alle Bundesländer verneint werden. Ausnahme bilden hier die Vergleich zu Bremen und Hessen. Eine Überschneidungsfreiheit beim Vergleich der ostdeutschen Vertrauensbereiche ist nicht gegeben.

Dass ein Arbeitnehmer in Ostdeutschland eine höhere Rendite erwarten darf, ist bereits aus Abschnitt 6.1 ersichtlich. Aufgrund des Vergleichs der Konfidenzintervalle kann nun

¹⁹⁴Die einwohnerstarken Bundesländer sind auch in der Stichprobe mit deutlich größerer Beobachtungszahl vertreten, sodass einzelne Abweichungen weniger ins Gewicht fallen.

mit diesem Verfahren davon ausgegangen werden, dass sich die hohe Bildungsrendite in Mecklenburg-Vorpommern tatsächlich von fast allen in Westdeutschland unterscheidet. Die Renditen aus den anderen Bundesländern können jedoch nicht als signifikant verschieden zueinander deklariert werden.

Um eine statistische Absicherung zu erhalten, wird im nachfolgenden Abschnitt ein zusätzliches Regressionsmodell geschätzt. Dabei wird die Stichprobe nicht separiert, sondern die regionale Variation wird über Interaktionsterme in der Regressionsgleichung berücksichtigt. Der Vorteil dabei ist, dass zum einen der Stichprobenumfang genügend groß ist sowie ein statistischer Test für Gleichheit der geschätzten Regressionskoeffizienten durchgeführt werden kann.

6.2.2 Regionale Bildungsrenditen auf Basis einer gesamtdeutschen Stichprobe

Zunächst wird in diesem Abschnitt das neue Regressionsmodell näher spezifiziert und beschrieben. Im nächsten Schritt folgt eine Diskussion über die zu verwendende Methode, um dann im dritten Schritt die Ergebnisse darzustellen. Abschließend wird ein Vergleich der mit diesem Modell geschätzten regionalen Bildungsrenditen auf Bundesländerebene gezeigt.

Modellbeschreibung. Um die unterschiedlichen Bildungsrenditen durch statistische Tests abzusichern, wird ein neues Modell zur Schätzung herangezogen. Dabei gilt

$$\begin{aligned} \ln Y_i = & \beta_0 + \sum_{l=1}^{15} \beta_l (D_{BL} \cdot S)_{il} + \beta_{16} Erf_i + \beta_{17} Erf_i^2 + \beta_{18} weibl_i + \beta_{19} verhei_i \\ & + \beta_{20} IATW-verh_i + \beta_{21} D_{gro\beta}_i + \beta_{22} D_{l\ddot{a}ndl}_i + \beta_{23} D_{pendler}_i + \sum_{l=24}^{37} \beta_l D_{BL-il} + \varepsilon_i, \end{aligned}$$

wobei die neue Variable in Tabelle 6.13 beschrieben ist.

Bezeichnung	Variablenname	Skalierung	vermuteter Zusammenhang
$D_{BL_{il}}$	Bundesland	dichotom, (1: Wohnhaft im l-ten Bundesland, 0: sonst)	Im IAT: positiv; Als Dummy: relativ zu SH zu bewerten.

Tabelle 6.13: Variablenbezeichnung – Interaktionsterme.

Im Unterschied zur vorhergehenden Analyse wird nun für jedes Bundesland ein Interaktionsterm in das Modell aufgenommen. Dabei steht $D_{BL_{il}}$ für eine Dummy-Variable mit der Ausprägung 1 in Bundesland¹⁹⁵ l , wobei $l = 1..15$. Der Interaktionsterm im l -ten Bundesland, berechnet durch: $\sum_{l=1}^{15} (D_{BL}S)_{il}$ und erhält lediglich dann die tatsächliche Ausprägung in der Höhe der individuellen Bildungsdauer, wenn das Wirtschaftssubjekt im l -ten Bundesland wohnt. Wohnt das Wirtschaftssubjekt nicht im l -ten Bundesland, so hat der Interaktionsterm die Ausprägung null. Im Unterschied zur bisherigen Analyse werden nun damit die einzelnen regionalen Bildungsrenditen innerhalb dieses Modells für eine gesamtdeutsche Stichprobe geschätzt.

Aus den vorangegangenen Abschnitten ist ein deutlicher Niveauunterschied im Einkommen pro Bundesland erkennbar. Damit diese Einflüsse nicht zu einer Verzerrung der Renditen führen, wird zusätzlich mit je einer Dummy-Variablen pro Bundesland für die unterschiedlichen Gegebenheiten kontrolliert. Dabei werden lediglich 14 Bundesländer berücksichtigt und der sich ergebende Koeffizient ist dann relativ zu einem Referenzbundesland zu interpretieren. Als Vergleichsbundesland wird Schleswig-Holstein¹⁹⁶ gewählt.

Verfahren. In den Abschnitten 5.1 und 5.2 wurde als sinnvolles Verfahren zur Schätzung der Bildungsrendite das 3SLS-Verfahren identifiziert. Darauf aufbauend ergeben sich alle bisherigen Regressionskoeffizienten. Für diese gesamtdeutsche Modellspezifikation kann jedoch nicht das 3SLS-Verfahren verwendet werden. In Abschnitt 5.2.1 wird in Annahme IV – V4 auf den Rang der Reduced-Form-Regression eingegangen. Dabei heißt es, dass dieser größer oder gleich der Schätzgleichung für den Lohn sein muss. Konkret bedeutet dies, dass es mindestens so viele Instrumente zur Schätzung der Bildungsdauer geben muss wie endogene Variablen in der Lohngleichung vorhanden sind. Bisher war diese Annahme problemlos erfüllt. Es gab mit der Bildungsdauer lediglich eine endogene Variable im Modell. In der neuen Modellspezifikation wird jedoch pro Bundesland eine

¹⁹⁵Durch die Aggregation von Saarland und Rheinland-Pfalz ist es ein Bundesland weniger als normalerweise.

¹⁹⁶Welches Bundesland hierfür gewählt wird, ist irrelevant.

Bildungsdauer berücksichtigt. Demnach ergeben sich 15 endogene Variablen. Die Zahl der Instrumente lässt sich jedoch durch die geringe Verfügbarkeit geeigneter Instrumente und die Notwendigkeit der sinnvollen Auswahl dieser nicht beliebig erweitern. Demzufolge ist dieses Modell unteridentifiziert und demnach mit dem 3SLS-Verfahren nicht schätzbar. Um trotzdem die Erkenntnisse aus der gemeinschaftlichen Regression zu erhalten, wird dieses Modell mit dem OLS-Verfahren geschätzt. Damit sind Verzerrungen der geschätzten Bildungsrenditen aufgrund des Endogenitätsproblems nicht auszuschließen. Die Interpretation der Höhe der Koeffizienten ist daher nur eingeschränkt möglich. Der Vergleich zwischen den Bundesländern ist, da die Verzerrung in gleicher Weise wirkt, näherungsweise jedoch möglich. Nachfolgend werden die geschätzten regionalen Bildungsrenditen dargestellt und diskutiert.

Ergebnisdarstellung. Die Ergebnisse aus der OLS-Regression mit Interaktionstermen werden in Tabelle 6.14 gezeigt. Dabei werden in den ersten beiden Spalten der Tabelle die aus den Interaktionstermen ermittelten Bildungsrenditen pro Bundesland dargestellt. Der zweite Teil zeigt die Koeffizienten der Kontrollvariablen, die bereits in den vorangegangenen Modellspezifikationen verwendet wurden. In den letzten beiden Spalten werden die Koeffizienten zu den Bundesländer-Dummy-Variablen dargestellt. Diese müssen relativ zum Referenzbundesland Schleswig-Holstein interpretiert werden.



Abbildung 6.6: Bildungsrenditen pro Bundesland (IAT); Quelle: DIW, eigene Berechnungen.

In Abbildung 6.6 werden die verschiedenen Bildungsrenditen anhand einer Deutschlandkarte veranschaulicht. Die geschätzten Renditen mit dieser Spezifikation liegen zwischen 6,8 %¹⁹⁷ in Hamburg und 15,3 % in Mecklenburg-Vorpommern. Auch die Rendite für Hamburg kann nun signifikant geschätzt werden. Die in Abschnitt 6.1 bereits ermittelte höhere Bildungsrendite in den ostdeutschen Bundesländern ist auch mit dieser Herangehensweise nachweisbar. Die niedrigste Rendite in den ostdeutschen Ländern wird in Brandenburg mit 9,4 % gemessen. In gleicher Höhe liegt die höchste Rendite der westdeutschen Bundesländer. Sie wird für Bremen gemessen. Anhand der Abbildung 6.6 sind nicht nur im Ost-West-Vergleich Unterschiede in den Bildungsrenditen zu erkennen. Auch in Ostdeutschland können verschiedene Renditen gemessen werden. So wird deutlich, dass in Mecklenburg-Vorpommern, Sachsen-Anhalt und Thüringen tendenziell höhere Renditen berechenbar sind als in Berlin, Brandenburg und Sachsen. Die Koeffizienten der Kontrollvariablen sind fast alle signifikant und entsprechen in Betrag und Richtung des Zusammenhangs den Erwartungen. Einzig der Schätzer für das ländliche Gebiet kann nicht signifikant nachgewiesen werden.

Durch die Dummy-Variablen für die Bundesländer werden generelle Niveauunterschiede im Einkommen berücksichtigt. Die Koeffizienten sind relativ zu Schleswig-Holstein zu interpretieren. Auffällig dabei ist, dass für alle westdeutschen Bundesländer kein signifikanter Niveauunterschied zum Referenzbundesland nachweisbar ist. Für alle ostdeutschen Bundesländer dagegen sind die Koeffizienten mit einer Irrtumswahrscheinlichkeit von 5 % signifikant von null verschieden. Da diese Koeffizienten alle ein negatives Vorzeichen haben, ist von einem generell geringeren Einkommen in den Bundesländern in Ostdeutschland im Vergleich zu Schleswig-Holstein auszugehen. Dabei kann für Brandenburg die geringste, für Mecklenburg-Vorpommern die höchste Abweichung gezeigt werden.

Für die neue Modellspezifikation müssen die in Abschnitt 5.1 dargestellten Voraussetzungen des OLS-Verfahrens neu geprüft werden. Generell können mit diesem Modell 42 % der gesamten Streuung erklärt werden. Der F-Test auf die Gesamtsignifikanz bestätigt auch für dieses Modell die Übertragbarkeit. Der Jarque-Bera-Test zeigt die Ablehnung der Annahme der normalverteilten Störgröße. Doch auch in diesem Modell ist der Stichprobenumfang größer als 60 Beobachtungen, sodass für die Koeffizienten auch in dieser Spezifikation näherungsweise von einer Normalverteilung auszugehen ist. Da die Annahme der Homoskedastizität über den Breusch-Pagan-Test abzulehnen ist, sind in der Ergebnistabelle 6.14 die korrigierten Standardfehler nach White angegeben.

¹⁹⁷Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs wie folgt berechnet:

$$\text{prozentuale Bildungsrendite} = (e^{\beta_1} - 1) \cdot 100.$$

6 Regionale Bildungsrenditen

Interaktions-terme	Koeffi- zienten	Kontroll- variablen	Koeffi- zienten	Bundes- länder- Dummys	Koeffi- zienten
BW	0,082*** [0,005]	Erfahrung	0,029*** [0,003]	BW	-0,053 [0,140]
BY	0,081*** [0,006]	Erfahrung ²	-0,000*** [0,000]	BY	-0,115 [0,147]
BE	0,102*** [0,012]	weiblich	-0,133*** [0,019]	BE	-0,533** [0,220]
BB	0,090*** [0,014]	verheiratet	0,096*** [0,017]	BB	-0,4724** [0,223]
HB	0,090*** [0,013]	IAT verh. Frau	-0,123*** [0,023]	HB	-0,338 [0,241]
HH	0,067*** [0,016]	großes Unternehmen	0,195*** [0,011]	HH	0,050 [0,280]
HE	0,082*** [0,007]	ländliches Gebiet	-0,022* [0,011]	HE	-0,094 [0,154]
MV	0,142*** [0,019]	Pendler	-0,048*** [0,011]	MV	-1,169*** [0,270]
NI	0,076*** [0,007]	Konstante	1,472*** [0,139]	NI	-0,077 [0,158]
NW	0,086*** [0,005]			NW	-0,171 [0,141]
RP / SL	0,086*** [0,007]			RP / SL	-0,202 [0,157]
SN	0,091*** [0,008]			SN	-0,538*** [0,168]
ST	0,114*** [0,012]			ST	-0,846*** [0,206]
SH	0,075*** [0,010]			TH	-0,871*** [0,187]
TH	0,115*** [0,010]				
R^2			0,42		
F-Test			0,000		
JB-Test			0,000		
BP-Test			0,000		
n			4.693		

Tabelle 6.14: Ergebnistabelle der OLS-Regression Interaktionstermen; es sind die Standardfehler nach White mit den Signifikanzniveaus: *** 1 %, ** 5 % und * 10 % angegeben. Die Testergebnisse werden über die P-Werte gezeigt. Quelle: DIW, eigene Berechnungen.

Der Durbin-Wu-Hausmann-Test ist nicht durchführbar, da die IV-Methode als Vergleichsinstrument aufgrund des geringeren Rangs der Reduced-Form-Regression und dem damit unteridentifizierten Modell nicht zur Verfügung steht. Es ist jedoch davon auszugehen, dass die Interaktionsterme der Bildungsrenditen ebenso wie beim ursprünglichen Modell endogen sind und Verzerrungen verursachen können. Räumliche Autokorrelation ist in diesem Modell nicht zu beachten, da die Bundesländer regional separiert sind und demzufolge die Koeffizienten den regionalen Aspekt mit berücksichtigen. Problematisch ist die Multikollinearität. Die Varianzinflationsfaktoren sind im Anhang A.4.2 in Tabelle A.19 dargestellt und übersteigen zum Teil den kritischen Wert von 20. Für die entscheidenden Interaktionsterme zur Berechnung der Bildungsrenditen muss demnach von einem Zusammenhang mit anderen erklärenden Variablen ausgegangen werden. Allerdings hält sich die Überschreitung mit maximal bis zu einem Wert von 30 in Grenzen, sodass dieses Problem vorhanden ist, der verzerrende Einfluss jedoch nicht als übermäßig bewertet werden muss.

Zusammenfassend kann festgestellt werden, dass Verzerrungen durch dieses Verfahren nicht ausgeschlossen sind. Wenn man jedoch die Ergebnisse im Vergleich zu den 3SLS-Koeffizienten betrachtet, können deutliche Gemeinsamkeiten in den Bildungsrenditen festgestellt werden. Da sich die Relationen zwischen den Bundesländern mit beiden Verfahren ähneln, können die Ergebnisse als Vergleichsindikator herangezogen werden.

Im nächsten Schritt werden die 95 %-Konfidenzintervalle gegenübergestellt, um einen besseren Eindruck über die Schätzergebnisse für die Bildungsrenditen über Interaktionsterme zu erhalten.

Bundesland	Bildungsrendite	Konfidenzintervall	Bundesland	Bildungsrendite	Konfidenzintervall
BW	0,082***	0,073–0,091	NI	0,076***	0,061–0,090
BY	0,081***	0,069–0,093	NW	0,086***	0,077–0,095
BE	0,102***	0,079–0,125	RP / SL	0,086***	0,072–0,100
BB	0,090***	0,063–0,117	SN	0,091***	0,074–0,108
HB	0,090***	0,064–0,116	ST	0,114***	0,091–0,137
HH	0,067***	0,036–0,097	SH	0,075***	0,056–0,094
HE	0,082***	0,069–0,096	TH	0,115***	0,095–0,135
MV	0,142***	0,106–0,179			

Tabelle 6.16: 95 %-Konfidenzintervalle aus der Regression mit Interaktionstermen für die Bundesländer; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; Quelle: DIW, eigene Berechnungen.

Regionaler Vergleich der Bildungsrenditen über Konfidenzintervalle. In Tabelle 6.7 sind die sich ergebenden Konfidenzintervalle pro Bundesland aus der neuen Modellspezifikation zusammengefasst. Alle Koeffizienten können mit dieser Spezifikation signifikant geschätzt werden. In Abbildung 6.7 werden die Konfidenzintervalle anhand eines Zahlenstrahls verdeutlicht. Wie in Abbildung 6.5 sind auch hier die Intervalle der ostdeutschen Bundesländer oberhalb, die der westdeutschen Bundesländer unterhalb des Zahlenstrahls dargestellt.

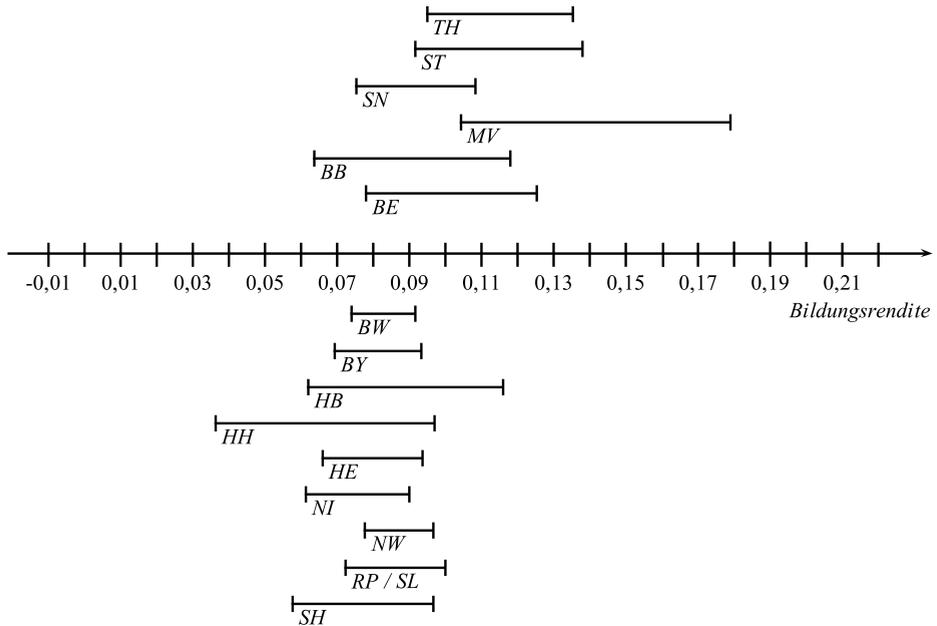


Abbildung 6.7: 95 %-Konfidenzintervalle für die Bildungsrenditen pro Bundesland (IAT); Quelle: DIW, eigene Darstellung.

Im Vergleich zu den Intervallen aus dem 3SLS-Verfahren sind die Spannweiten dieser Konfidenzintervalle deutlich geringer. Jedoch kann auch für diesen Vergleich die Überschneidungsfreiheit der Intervalle nur zwischen den westdeutschen Bundesländern – mit Ausnahme von Bremen – mit Mecklenburg-Vorpommern gezeigt werden. Die Intervalle der ostdeutschen Bundesländer überschneiden sich komplett. Demnach kann auch aus diesem Modell eine Verschiedenartigkeit zwischen Mecklenburg-Vorpommern und den westdeutschen Bundesländern festgestellt werden. Keine Unterschiede zeigt der intrare-

gionale Vergleich der westdeutschen oder ostdeutschen Bundesländer. Selbst zwischen den Ost- und West-Ländern können ansonsten keine Unterschiede garantiert werden.

Regionaler Vergleich der Bildungsrenditen über den Wald-Test. Werden Koeffizienten innerhalb *eines* Regressionsmodells geschätzt, können diese über den Wald-Test¹⁹⁸ miteinander verglichen werden. Dabei wird folgende Nullhypothese getestet:

$$H_0 : \beta_i - \beta_j = 0$$

$$H_1 : \beta_i - \beta_j \neq 0,$$

wobei $i \neq j$ gilt. Kann die Nullhypothese mit einer Irrtumswahrscheinlichkeit von 5 % abgelehnt werden, ist davon auszugehen, dass sich die Koeffizienten der Regression signifikant unterscheiden. In diesem Abschnitt wird jede Bildungsrendite pro Bundesland mit den jeweils anderen Bundesländern verglichen. Die sich ergebenden P-Werte sind in Tabelle 6.17 dargestellt. Zugrunde liegt eine F-Verteilung.¹⁹⁹ Ist eine signifikante Ablehnung der Nullhypothese möglich, werden die Werte in der Tabelle kursiv angegeben. Aus dem Wald-Test ergibt sich wiederum die deutliche Ablehnung der Nullhypothese für die Bildungsrenditen im Vergleich zu Mecklenburg-Vorpommern. Durch den statistischen Test können jedoch auch einzelne Bildungsrenditen ostdeutscher Bundesländer im Vergleich zu Mecklenburg-Vorpommern abgelehnt werden. Somit sind signifikante Unterschiede zwischen diesem Bundesland und Brandenburg sowie Sachsen nachweisbar. Erweitert man die Irrtumswahrscheinlichkeit auf 10 %, kann auch ein signifikanter Unterschied zu Berlin festgestellt werden.

Abgesehen von Mecklenburg-Vorpommern ergibt sich auch für den Vergleich mit anderen Bundesländern die Möglichkeit der Ablehnbarkeit der Nullhypothese. Mit einer Sicherheit von 95 % kann die Gleichheit der Bildungsrendite für den Vergleich von Sachsen-Anhalt mit Baden-Württemberg, Bayern, Hamburg, Hessen, Niedersachsen, Nordrhein-

¹⁹⁸Nähere Beschreibung und ökonomische Herleitung des Tests siehe Heij et al. (2004), S. 232 ff und 240 ff. Dieser Test ist nicht zu verwechseln mit dem Chow-Test. Dabei werden *alle* Koeffizienten eines Regressionsmodells zweier unabhängiger Stichproben gegeneinander getestet.

¹⁹⁹In Heij et al. (2004), S. 242, wird die beste Verteilung diskutiert. Üblicherweise folgt ein Wald-Test einer χ^2 -Verteilung. In diesem Fall liefert die F-Verteilung stabilere Ergebnisse, da die kritischen Werte der ersten Verteilung niedriger sind. Allgemein jedoch nähern sich beide Verteilungen für große Stichprobenumfänge einander an.

Wald-Test	BW	BY	BE	BB	HB	HH	HE	MV	NI	NW	PR / SL	SN	ST	SH	TH
BW	–														
BY	0,90	–													
BE	0,12	0,12	–												
BB	0,58	0,54	0,52	–											
HB	0,56	0,53	0,52	1,00	–										
HH	0,34	0,39	0,07	0,26	0,25	–									
HE	0,98	0,90	0,15	0,60	0,59	0,36	–								
MV	0,00	0,00	0,07	0,02	0,02	0,00	0,00	–							
NI	0,46	0,57	0,06	0,35	0,34	0,60	0,51	0,00	–						
NW	0,52	0,50	0,22	0,78	0,77	0,23	0,63	0,00	0,22	–					
RP / SL	0,67	0,63	0,24	0,77	0,76	0,27	0,73	0,00	0,33	0,95	–				
SN	0,34	0,22	0,47	0,95	0,95	0,16	0,41	0,01	0,17	0,60	0,62	–			
ST	0,01	0,01	0,46	0,18	0,18	0,01	0,02	0,20	0,01	0,03	0,04	0,11	–		
SH	0,51	0,60	0,40	0,37	0,36	0,64	0,55	0,00	0,97	0,30	0,38	0,21	0,01	–	
TH	0,00	0,00	0,40	0,15	0,14	0,01	0,01	0,20	0,00	0,01	0,02	0,95	0,95	0,00	–

Tabelle 6.17: Ergebnisse des Wald-Tests auf Unterschriedlichkeit der Bildungsrenditen pro Bundesland; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; die kursiv geschriebenen Zahlen lassen die Ablehnung der Gleichheitshypothese zu. Quelle: DIW, eigene Berechnungen.

Westfalen, Rheinland-Pfalz und Saarland sowie Schleswig-Holstein jeweils abgelehnt werden. Gleiches gilt für den Vergleich der Bildungsrenditen mit dem Ergebnis aus Thüringen.

Alle weiteren Vergleiche ergeben mit einer Irrtumswahrscheinlichkeit von 5 % keine Möglichkeit der Ablehnung der Gleichheitshypothese. Diese Herangehensweise zeigt zum einen eine Bestätigung der Erwartungen, die durch die Konfidenzintervalle aufgestellt wurden. Zudem können auch weitere Renditen als signifikant verschieden deklariert werden. Im Besonderen fallen nun nach dem Vergleich die Bundesländer Mecklenburg-Vorpommern, Sachsen-Anhalt und Thüringen auf. Für alle anderen Bundesländer ergibt sich keine empirische Evidenz für die Unterschiedlichkeit in den Bildungsrenditen.

6.2.3 Vergleich der regionalen Bildungsrenditen

In diesem Abschnitt werden die verschieden geschätzten regionalen Bildungsrenditen miteinander verglichen. Wie bereits dargestellt sind Bildungsrenditen, die über das 3SLS-Verfahren berechnet werden, tendenziell höher als jene aus dem OLS-Verfahren. Daher werden als Vergleichsinstrument Bildungsrenditen aus dem nach Bundesländern separierten Datensatz nun zusätzlich über das OLS-Verfahren berechnet. Aus Gründen der Übersichtlichkeit sind die detaillierten Ergebnisse aus diesen Regressionen im Anhang A.4.3 in den Tabellen A.20 und A.21 gezeigt. In diesem Abschnitt ist aus dieser Regression lediglich die zusammenfassende Deutschlandkarte dargestellt. Für den Vergleich werden nun die geschätzten regionalen Bildungsrenditen in Tabelle 6.18 zusammengefasst.

Zu beachten ist dabei, dass die 3SLS-Methode (Tabelle 6.18, Spalte 2) und die OLS-Methode (Tabelle 6.18, Spalte 3) basierend auf dem separierten Datensatz mit der gleichen Schätzgleichung²⁰⁰ für den Lohn regressiert werden. Der Unterschied ergibt sich lediglich in einer Zusatzregression der Bildungsjahre, um dem Endogenitätsproblem zu begegnen. Mit einer anderen Schätzgleichung, auf Basis von Interaktionstermen, sind die Bildungsrenditen in Spalte 4 geschätzt.²⁰¹ Durch die verschiedenen Methoden²⁰² sind die Werte der OLS-Regressionen nur schwer mit denen des 3SLS-Verfahrens vergleichbar. Bei den OLS-Regressionen ist für beide Modellvarianten von einem Endogenitätsproblem auszugehen. Im Hinblick auf die bisherige Analyse ist tendenziell zu erwarten, dass die Ergebnisse aus der 3SLS-Regression höher sind als die der beiden anderen. Diese Erwartung

²⁰⁰Siehe Abschnitt 5.1.1.

²⁰¹Siehe Abschnitt 5.2.1.

²⁰²Im Detail wurden die Herangehensweisen beider Verfahren in den Abschnitten 5.1 (OLS) und 5.2 (3SLS) gezeigt.

6 Regionale Bildungsrenditen

kann für fast alle Bundesländer bestätigt werden, einzig Thüringen und Hamburg bilden eine Ausnahme. Die Änderungen in Thüringen sind jedoch so gering, dass sie vernachlässigbar sind. Die Hamburger Bildungsrendite ist beim 3SLS-Verfahren nicht signifikant, sodass sie aus dem Vergleich herauszunehmen ist. Für die Bundesländer Berlin, Baden-Württemberg, Bayern, Rheinland-Pfalz und Saarland, Schleswig-Holstein und Thüringen können für alle drei Verfahren sehr ähnliche Bildungsrenditen geschätzt werden. Für die Bundesländer Brandenburg, Bremen, Hessen, Mecklenburg-Vorpommern, Niedersachsen, Hamburg und Nordrhein-Westfalen sind zumindest bei beiden OLS-Verfahren, unabhängig vom gewählten Regressionsmodell, sehr ähnliche Ergebnisse zu erhalten. Einzig für Sachsen und Sachsen-Anhalt ergeben sich unterschiedliche Bildungsrenditen je nach Modellwahl.

Bundesland	3SLS sep. Datensatz	OLS sep. Datensatz	OLS gesamtdeutscher Datensatz	Bundesland	3SLS sep. Datensatz	OLS sep. Datensatz	OLS gesamtdeutscher Datensatz
BW	0,088 (***)	0,080 (***)	0,082 (***)	NI	0,093 (***)	0,078 (***)	0,076 (***)
BY	0,087 (***)	0,084 (***)	0,081 (***)	NW	0,106 (***)	0,089 (***)	0,086 (***)
BE	0,106 (***)	0,100 (***)	0,102 (***)	RP / SL	0,092 (***)	0,085 (***)	0,086 (***)
BB	0,156 (***)	0,088 (***)	0,090 (***)	SN	0,137 (***)	0,084 (***)	0,091 (***)
HB	0,124 (***)	0,093 (***)	0,090 (***)	ST	0,138 (***)	0,107 (***)	0,114 (***)
HH	0,042 (*)	0,071 (***)	0,067 (***)	SH	0,079 (***)	0,078 (***)	0,075 (***)
HE	0,107 (***)	0,079 (***)	0,082 (***)	TH	0,112 (***)	0,110 (***)	0,115 (***)
MV	0,172 (***)	0,142 (***)	0,142 (***)				

Tabelle 6.18: Regionale Bildungsrenditen im Vergleich; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; Quelle: DIW, eigene Berechnungen.

Zum Vergleich werden diese Bildungsrenditen zusätzlich über Deutschlandkarten in Abbildung 6.8 veranschaulicht. Zur besseren Vergleichbarkeit sind alle Intervalle zwischen 6,5 % und 19,5 % gewählt. Wie auch aus Tabelle 6.18 ersichtlich, sind die Unterschiede in der Höhe der Bildungsrendite der beiden OLS-Modellspezifikationen pro Bundesland

gering. Dieses Ergebnis beweist die Stabilität der geschätzten Bildungsrenditen auch über verschiedene Modelle hinweg. Problematisch bleibt die fehlende Korrektur der Endogenität in dieser Schätzung.

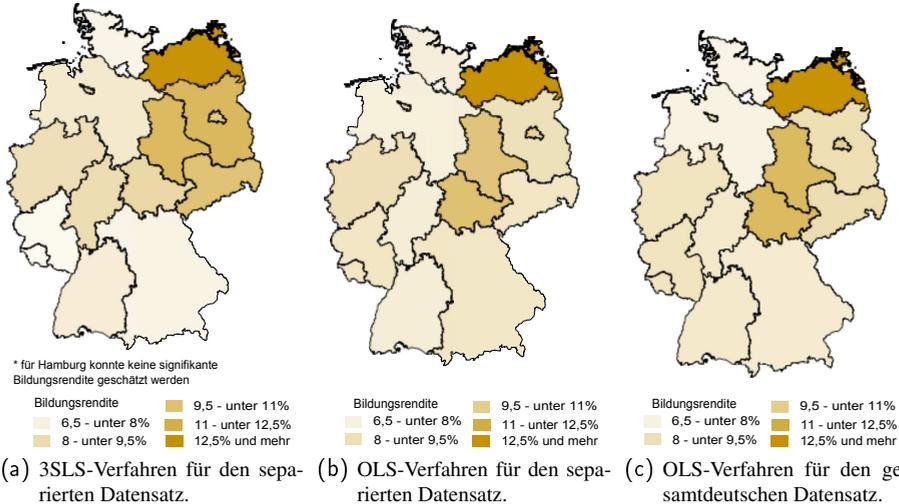


Abbildung 6.8: Regionale Bildungsrenditen im Vergleich; Quelle: DIW, eigene Berechnungen.

Betrachtet man ausschließlich die Relationen zwischen den Bundesländern im Vergleich zwischen OLS-Verfahren und 3SLS-Verfahren, zeigt sich Konsistenz. Über den Wald-Test in Abschnitt 6.2.2 sind signifikante Unterschiede in der Höhe der Bildungsrenditen für den Vergleich zwischen den Renditen von Mecklenburg-Vorpommern, Thüringen und Sachsen-Anhalt mit allen westdeutschen Bundesländern nachweisbar. Zum Teil ergibt sich auch die signifikante Ablehnung zu Berlin und Brandenburg sowie Sachsen. Auch wenn der statistische Test für das 3SLS-Verfahren nicht durchführbar ist, so zeigt sich doch die gleiche Tendenz, sodass auch hierbei Mecklenburg und Sachsen-Anhalt im Vergleich hervorzuheben sind.

6.2.4 Kritik

Die durchgeführte regionale Separierung basiert auf geschichtlich gewachsenen Grenzen. Dies muss nicht zwingend mit wirtschaftlich homogenen Räumen einhergehen. Im Ost-West-Vergleich ist jedoch nicht davon auszugehen, dass sich die Grenzziehung negativ auf die Ergebnisse auswirkt, da sich die Regionen bis zum Beobachtungsjahr auch wirtschaft-

lich unterscheiden. Anders ist es im intraregionalen Vergleich zwischen den Bundesländern, weil hier zum Beispiel die Regionen um die Stadtstaaten schwer von den angrenzenden Bundesländern zu trennen sind.

Die Analyse auf Bundesländerebene findet auf mehreren Wegen statt. Um auf der einen Seite einen Vergleichstest anwenden zu können, wird auf der anderen Seite das Endogenitätsproblem akzeptiert. Dieser Zielkonflikt muss bei der Interpretation der Koeffizienten berücksichtigt werden. Da jedoch die Relationen der Koeffizienten zwischen den Bundesländern in allen Modellspezifikationen und Verfahren stabil sind, können die Gemeinsamkeiten und Unterschiede in der Höhe der Bildungsrenditen in Deutschland herausgearbeitet werden.

6.2.5 Zusammenfassung

In diesem Abschnitt werden Bildungsrenditen auf Bundesländerebene berechnet. Die Ermittlung dieser aus einem nach den Bundesländern separierten Datensatz mit der 3SLS-Methode bestätigt die Ergebnisse auf Basis der Ost-West-Teilung. Aufgrund von statistischen Gegebenheiten kann die Höhe der Bildungsrenditen im Vergleich nicht auf Signifikanz getestet werden. Von der Höhe her fällt jedoch Mecklenburg-Vorpommern mit einer Bildungsrendite von 18,6 % pro zusätzlichem Bildungsjahr besonders auf. Zur Prüfung, ob dieses Bundesland sich nicht nur zufällig abhebt, wird ein neues Regressionsmodell geschätzt. Hierbei kommt eine gesamtdeutsche Stichprobe zum Einsatz, für die statistische Tests über den Wald-Test die Prüfung auf signifikante Gleichheit der Koeffizienten durchführbar ist. Problematisch ist, dass dieses Verfahren lediglich mit dem OLS-Verfahren geschätzt werden kann. Auch aus dieser Art der Berechnung regionaler Bildungsrenditen hebt sich Mecklenburg-Vorpommern deutlich hervor. Zusätzlich können auch im Vergleich die Bildungsrenditen von Sachsen-Anhalt und Thüringen häufig als signifikant verschieden von anderen Bundesländern getestet werden.

Der Vergleich im letzten Abschnitt zeigt in der Relation zwischen den Bundesländern konsistente Ergebnisse, unabhängig von Modellspezifikation und Methode. Deutschland ist bezüglich der Bildungsrenditen für die meisten Bundesländer als durchaus homogen anzusehen. Es ergibt sich eine leichte Tendenz zu einer höheren mittleren jährlichen Verzinsung in den neuen Bundesländern. Deutlich verschiedenartig ist Mecklenburg-Vorpommern. Beim Vergleich der Teststatistiken ergibt sich auch eine Abgrenzbarkeit für Sachsen-Anhalt und Thüringen.

6.3 Regionale Bildungsrenditen unter Berücksichtigung der Erwerbslosen

Analog zum Abschnitt 6.2.1 werden an dieser Stelle die einzelnen Bundesländer separat analysiert. Dabei wird zunächst das Heckman-Verfahren für einen nach Bundesländern separierten Datensatz angewendet. Im nächsten Schritt wird wiederum zur besseren Vergleichbarkeit der Koeffizienten eine Regression für den gesamten Datensatz mit Variation der Bildungsdauern über Interaktionsterme geschätzt. Aufgrund des bereits dargestellten Verfahrens in Abschnitt 5.4.3 sowie der bereits durchgeführten Anwendung der Regressionsmodelle für das 3SLS-Verfahren sind an dieser Stelle lediglich die neuen Ergebnisse dargestellt.

6.3.1 Regionale Bildungrenditen auf Basis separierter Stichproben

Die Modellspezifikation für die vorgelagerte Probit-Regression ist dargestellt in Abschnitt 5.4.3, das Lohngleichungsmodell entspricht dem bereits in Abschnitt 5.1.1 identifizierten Modell 4. In den Tabellen 6.19 und 6.20 sind die Ergebnisse aus der Lohngleichungsregression des Heckman-Verfahrens separiert pro Bundesland dargestellt. Die jeweiligen Koeffizienten der vorgelagerten Probit-Regressionen pro Bundesland sind im Abschnitt A.4.4 gezeigt.

Zusätzlich wird die jeweilige Bildungsrendite pro Bundesland in Abbildung 6.9 zusammengefasst. Zur besseren Vergleichbarkeit ist das Intervall zur Einfärbung der Karte in gleicher Weise gewählt wie bei den vorherigen Deutschlandkarten in Abschnitt 6.2.1. Für Bremen und Hamburg kann durch die Separierung die Heckman-Methode nicht angewendet werden. Bei dem Verfahren ist das Ziel, einen Korrekturvektor für die verschiedenen Eigenschaften zu ermitteln. Da jedoch bei den Stichproben der beiden Bundesländer nur kleine Beobachtungszahlen verfügbar sind, ist die Ermittlung dieses Korrekturfaktors hierbei nicht möglich. Im separierten Datensatz bleiben in Hamburg 7 und in Bremen lediglich 2 Erwerbslose berücksichtigt. Mit diesen Fallzahlen ist keine statistische Analyse möglich. Alle weiteren Stichprobenumfänge sind deutlich größer und können für die Analyse verwendet werden.

Beim Vergleich der Renditen zeigt sich, dass die geringste prozentuale Verzinsung von allgemeinem Humankapital in Schleswig-Holstein berechnet werden kann. Hier verdient

6 Regionale Bildungsrenditen

im Mittel eine Erwerbsperson zwischen dem 30. und 60. Lebensjahr pro zusätzlichem Bildungsjahr 7,4 %²⁰³ mehr. Die höchste Bildungsrendite ist auch mit diesem Verfahren für Mecklenburg-Vorpommern berechnet worden. Dabei liegt die mittlere Verzinsung von Bildung bei 13,4 % pro Jahr. Unterscheidet man in die Ost- und West-Region, so zeigt sich, dass die höchste Bildungsrendite in den westdeutschen Bundesländern mit diesem Verfahren in Bayern (8,4 %) zu ermitteln ist. Im Gegenzug verzinst sich das allgemeine Humankapital in Sachsen mit 7,6 % im Vergleich zu den weiteren ostdeutschen Bundesländern am wenigsten.



Abbildung 6.9: Grafische Ergebnisdarstellung der nach Bundesländern separierten Heckman-Regression; Quelle: DIW, eigene Berechnungen.

Beim Vergleich der Ergebnisse – bezogen auf die Kontrollvariablen – ist festzustellen, dass für die signifikanten Koeffizienten bei allen Teilregressionen der erwartete Zusammenhang nachgewiesen werden kann. Aufgrund der reduzierten Datensätze sind jedoch viele Koeffizienten nicht signifikant. Bemerkenswert ist, dass der Einfluss der Unterneh-

²⁰³Die prozentuale Zahl wird aufgrund des semilogarithmischen Zusammenhangs wie folgt berechnet:
$$\text{prozentuale Bildungsrendite} = (e^{\beta_1} - 1) \cdot 100.$$

6 Regionale Bildungsrenditen

Regressoren	BW	BY	BE	BB	HE	MV	NI
Bildungsjahre	0,076*** (0,005)	0,081*** (0,006)	0,085*** (0,017)	0,076*** (0,017)	0,076*** (0,007)	0,126*** (0,017)	0,076*** (0,008)
Erfahrung	0,033*** (0,008)	0,049*** (0,010)	0,057*** (0,021)	0,011 (0,023)	0,021* (0,011)	0,040 (0,029)	0,040*** (0,012)
Erfahrung ²	-0,001*** (0,000)	-0,001*** (0,000)	-0,001** (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,001 (0,001)	-0,001** (0,000)
weiblich	-0,144*** (0,047)	-0,142*** (0,053)	-0,246* (0,131)	-0,213 (0,132)	-0,100 (0,074)	-0,137 (0,136)	-0,190*** (0,072)
verheiratet	0,117*** (0,039)	0,063 (0,045)	-0,244* (0,132)	-0,089 (0,128)	0,047 (0,066)	0,016 (0,118)	0,185*** (0,057)
IAT verheiratete Frau	-0,156*** (0,056)	-0,128** (0,063)	0,205 (0,168)	0,066 (0,158)	-0,084 (0,085)	0,041 (0,155)	-0,063 (0,084)
großes Unternehmen	0,125*** (0,026)	0,137*** (0,030)	0,294*** (0,085)	0,198*** (0,075)	0,197*** (0,038)	0,123* (0,073)	0,155*** (0,039)
Pendler	0,061** (0,027)	0,027 (0,030)	0,096 (0,121)	0,136* (0,078)	0,122*** (0,039)	-0,069 (0,073)	0,020 (0,040)
ländliches Gebiet	-0,054** (0,027)	0,071** (0,030)	0,075 (0,149)	-0,048 (0,073)	-0,046 (0,037)	0,081 (0,074)	-0,049 (0,038)
Konstante	1,428*** (0,132)	1,104*** (0,158)	0,978*** (0,374)	1,640*** (0,376)	1,451*** (0,189)	0,316 (0,454)	1,109*** (0,194)
λ	-0,245*** (0,078)	-0,225* (0,127)	-0,387*** (0,140)	-0,537*** (0,128)	-0,164 (0,123)	-0,371*** (0,124)	-0,212* (0,122)
Wald-Test n (n Arbeitslos)	0,000 622 (34)	0,000 714 (37)	0,000 169 (36)	0,000 232 (46)	0,000 374 (25)	0,000 145 (36)	0,000 449 (37)

Tabelle 6.19: Ergebnistabelle der Heckman-Regression auf Bundesländerebene, Teil 1; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

6 Regionale Bildungsrenditen

Regressoren	NW	RP / SL	SN	ST	SH	TH
Bildungsjahre	0,080*** (0,005)	0,080*** (0,009)	0,073*** (0,009)	0,083*** (0,020)	0,071*** (0,011)	0,092*** (0,018)
Erfahrung	0,021** (0,008)	0,041*** (0,016)	0,019* (0,011)	0,050* (0,027)	0,047*** (0,016)	-0,000 (0,026)
Erfahrung ²	-0,000 (0,000)	-0,001** (0,000)	-0,000 (0,000)	-0,001* (0,000)	-0,001** (0,000)	-0,000 (0,001)
weiblich	-0,176*** (0,045)	-0,177** (0,082)	-0,031** (0,069)	-0,046 (0,163)	-0,115 (0,089)	-0,057 (0,124)
verheiratet	0,058 (0,038)	0,090 (0,069)	-0,015 (0,060)	0,144 (0,154)	0,033 (0,086)	0,132 (0,120)
IAT verheiratete Frau	-0,136*** (0,052)	-0,194** (0,094)	-0,107 (0,081)	-0,169 (0,191)	-0,087 (0,110)	-0,252 (0,154)
großes Unternehmen	0,177*** (0,024)	0,225*** (0,043)	0,245*** (0,040)	0,252*** (0,096)	0,131*** (0,050)	0,313*** (0,078)
Pendler	0,044* (0,024)	0,091** (0,044)	-0,029 (0,039)	0,003 (0,097)	0,019 (0,055)	0,021 (0,079)
ländliches Gebiet	-0,023 (0,028)	-0,052 (0,042)	-0,015 (0,038)	-0,025 (0,092)	-0,013 (0,054)	-0,038 (0,077)
Konstante	1,427*** (0,135)	1,125*** (0,250)	1,304*** (0,201)	0,849* (0,442)	1,236*** (0,262)	1,302*** (0,396)
λ	-0,352*** (0,064)	-0,283** (0,113)	-0,293*** (0,075)	-0,696*** (0,149)	0,286*** (0,094)	-0,589*** (0,130)
Wald-Test n (n Arbeitslos)	0,000 1,063 (92)	0,000 340 (30)	0,000 411 (62)	0,000 235 (40)	0,000 144 (16)	0,000 241 (5)

Tabelle 6.20: Ergebnistabelle der Heckman-Regression auf Bundesländerebene, Teil 2; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

mensgröße auch bei dieser Regression für fast alle²⁰⁴ Bundesländer zu signifikanten Ergebnissen führt. Eine Ausnahme bildet hierbei Mecklenburg-Vorpommern. Hier kann der Zusammenhang mit einer Irrtumswahrscheinlichkeit von 5 % nicht signifikant bestätigt werden.

Neu in dieser Regression ist der Korrekturterm. Mit Ausnahme von Bayern, Hessen und Niedersachsen kann dieser signifikant nachgewiesen werden und damit einhergehend wird auch das Selektionsproblem für diese Bundesländer bestätigt. Der Koeffizient hat für alle separierten Regressionen das erwartete negative Vorzeichen. Die sinnvolle Modellauswahl kann über den Wald-Test bestätigt werden.

Vergleich der 95 %-Konfidenzintervalle. Da auch hierfür bei der separierten Regression die Koeffizienten nicht über einen statistischen Test vergleichbar sind, werden die 95 %-Konfidenzintervalle als Indikatoren verwendet. Aus den berechneten Bildungsrenditen zeigt sich wenig Variabilität zwischen den Ergebnissen der westdeutschen Bundesländer.

Bundesland	Bildungsrendite	Konfidenzintervall	Bundesland	Bildungsrendite	Konfidenzintervall
BW	0,076***	0,066–0,086	NW	0,080***	0,070–0,090
BY	0,081***	0,070–0,093	RP / SL	0,080***	0,063–0,098
BE	0,085***	0,052–0,118	SN	0,073***	0,056–0,090
BB	0,076***	0,043–0,110	ST	0,083***	0,043–0,123
HE	0,076***	0,062–0,090	SH	0,071***	0,050–0,093
MV	0,126***	0,092–0,160	TH	0,092***	0,056–0,127
NI	0,076***	0,060–0,091			

Tabelle 6.22: 95 %-Konfidenzintervalle aus der Heckman-Regression separiert nach Bundesländern; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; Quelle: DIW, eigene Berechnungen.

In Tabelle 6.22 sind sie numerisch aufgezeigt, in Abbildung 6.10 werden sie dargestellt. Die Intervalle sind, wie auch bei der bisherigen Analyse, für den ostdeutschen Raum deutlich größer als in Westdeutschland. Die Überschneidungsfreiheit kann einzig für Mecklenburg-Vorpommern mit fast allen westdeutschen Bundesländern gezeigt werden. Eine Ausnahme ist Bayern, beide Intervalle überschneiden sich aber nur marginal. Zusätzlich kann die Überschneidungsfreiheit zwischen Mecklenburg-Vorpommern und Sachsen gezeigt werden. Alle weiteren Intervalle überschneiden sich gegenseitig. Demzufolge kann für

²⁰⁴Aufgrund der zu geringen Fallzahlen zur Anwendung der Heckmann-Methode beziehen die Interpretationen Hamburg und Bremen nicht mit ein.

diese nicht davon ausgegangen werden, dass sich die Renditen für diese Bundesländer tatsächlich unterscheiden. Dieser Vergleich der Konfidenzintervalle dient lediglich als Indikator und kann demnach nicht abschließend zur Beurteilung der Signifikanz herangezogen werden. Daher wird im nächsten Abschnitt das Heckman-Verfahren unter Verwendung von Interaktionstermen angewandt.

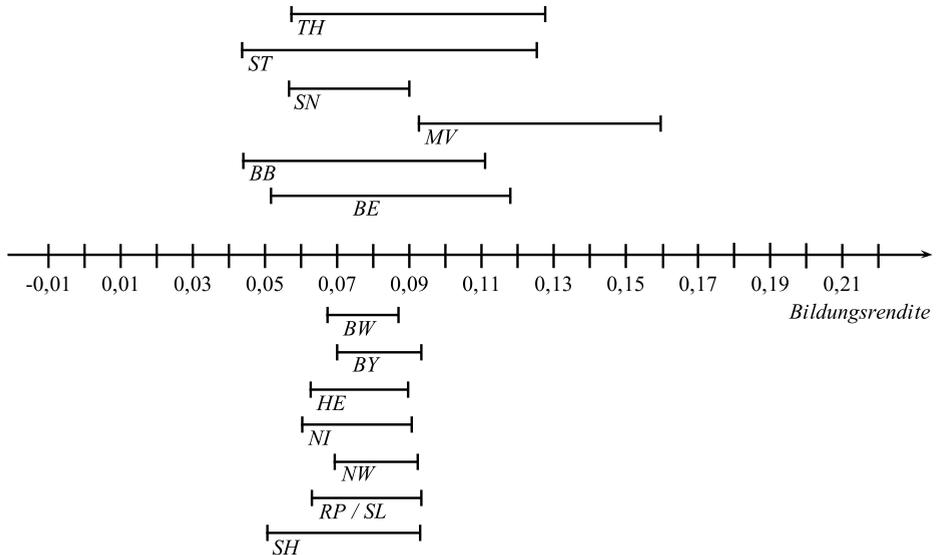


Abbildung 6.10: 95 %-Konfidenzintervalle der Bildungsrenditen mit der nach Bundesländern separierten Heckman-Methode; Quelle: DIW, eigene Darstellung.

6.3.2 Regionale Bildungsrenditen auf Basis einer gesamtdeutschen Stichprobe

Wie bereits in Abschnitt 6.2.1 wird das Modell umgestellt, sodass die Berechnung der regionalen Bildungsrenditen in einem gesamtdeutschen Datensatz vollzogen werden kann. Der Vorteil ist zum einen, dass die Datensatzreduzierungen nicht zu zu kleinen Stichprobenumfängen führen und zum anderen sind die Koeffizienten über den Wald-Test gegeneinander statistisch testbar. Im Unterschied zum oben genannten Abschnitt wird nun anstelle der OLS-Methode das Verfahren von Heckman verwendet. Die Modellspezifikation der vorgelagerten Probit-Regression entspricht der in Abschnitt 5.4.3 dargestellten. Die Schätzung der Lohngleichung erfolgt über die Spezifikation aus Abschnitt 6.2.2, wobei

6.3 Regionale Bildungsrenditen unter Berücksichtigung der Erwerbslosen

die regionale Variation über die Interaktionsterme berücksichtigt ist. Die Ergebnisse aus dieser Regression sind in Tabelle 6.23 dargestellt.

Interaktions-term	Koeffizienten	Kontrollvariablen	Koeffizienten	Bundesländer-Dummys	Koeffizienten
BW	0,076*** (0,005)	Erfahrung	0,029*** (0,003)	BW	-0,054 (0,177)
BY	0,074*** (0,005)	Erfahrung ²	-0,000*** (0,000)	BY	-0,111 (0,176)
BE	0,091*** (0,011)	weiblich	-0,124*** (0,020)	BE	-0,457** (0,226)
BB	0,086*** (0,011)	verheiratet	0,084*** (0,017)	BB	-0,493** (0,219)
HB	0,081*** (0,019)	IAT verh. Frau	-0,130*** (0,023)	HB	-0,285 (0,308)
HH	0,063*** (0,015)	großes Unternehmen	0,188*** (0,011)	HH	0,041 (0,277)
HE	0,077*** (0,007)	Pendler	0,042*** (0,011)	HE	-0,113 (0,188)
MV	0,135*** (0,015)	ländliches Gebiet	-0,021* (0,011)	MV	-1,150*** (0,254)
NI	0,070*** (0,007)	Konstante	1,495*** (0,171)	NI	-0,089 (0,186)
NW	0,079*** (0,004)	λ	-0,302*** (0,025)	NW	-0,165 (0,173)
RP / SL	0,080*** (0,008)			RP / SL	-0,206 (0,193)
SN	0,082*** (0,008)			SN	-0,491** (0,193)
ST	0,103*** (0,010)			ST	-0,768*** (0,211)
SH	0,069*** (0,013)			TH	-0,815*** (0,217)
TH	0,105*** (0,011)				
Wald-Test n (n Arbeitslos)	0,000 5.248 (555)				

Tabelle 6.23: Ergebnistabelle der Heckman-Regression mit IAT pro Bundesland; Quelle: DIW, eigene Berechnungen. Signifikanzniveau: *** 1%, ** 5% und * 10%.

In Abbildung 6.11 werden die unterschiedlichen Höhen der ermittelten regionalen Bildungsrenditen grafisch verdeutlicht. Tendenziell zeigen sich ähnliche Ergebnisse wie bei der separierten Heckman-Regression. Auch bei dieser Regression wurde für Mecklenburg-Vorpommern mit 14,5 % die mit Abstand höchste Bildungsrendite ermittelt. Die geringste Rendite konnte für Hamburg errechnet werden. Hierbei verzinst sich ein zusätzliches Bildungsjahr mit 6,5 %. Die höchste Rendite in Westdeutschland wurde für Rheinland-Pfalz und das Saarland mit 8,3 % berechnet. Im Vergleich der ostdeutschen Länder ist in Sachsen mit 8,5 % die niedrigste Rendite zu verzeichnen. Sie ist damit immer noch höher als die höchste der westdeutschen Bundesländer.

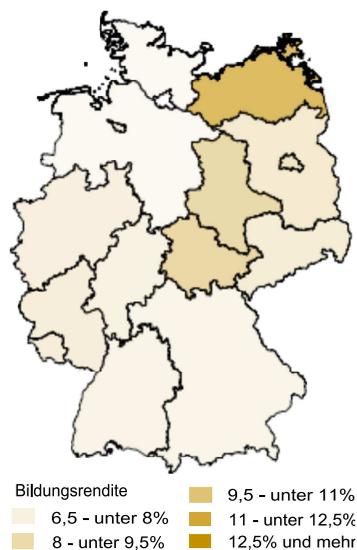


Abbildung 6.11: Grafische Ergebnisdarstellung der Bildungsrenditen aus der Heckman-Regression mit IAT; Quelle: DIW, eigene Berechnungen.

Alle weiteren Koeffizienten bestätigen ein weiteres Mal die Vermutungen über die Wirkungsweise auf das Einkommen. Die einzelnen Dummy-Variablen pro Bundesland werden auch an dieser Stelle hinzugefügt, um Verzerrungen durch grundsätzlich unterschiedliche Einkommensniveaus herauszurechnen. Sie müssen, wie bereits in Abschnitt 6.2.2 relativ zu Schleswig-Holstein²⁰⁵ beurteilt werden. Der Vergleich der Koeffizienten ergibt,

²⁰⁵Die Auswahl des Referenzbundeslandes ist willkürlich gewählt. Auch jedes andere Bundesland hätte demnach ohne Informationsverlust ausgewählt werden können.

dass die westdeutschen Bundesländer sich nicht grundlegend unterscheiden. Für keines dieser Bundesländer ist ein signifikanter Koeffizient nachweisbar. Anders hingegen fällt der Vergleich zu den ostdeutschen Bundesländern aus. Mit einer Irrtumswahrscheinlichkeit von 5 % kann für sie ein geringeres Einkommensniveau im Vergleich zu Schleswig-Holstein verzeichnet werden. Der Koeffizient des inversen Mills-Ratios zeigt erwartungsgemäß auch für diese Regression den signifikant negativen Zusammenhang. Demnach ist auch hierbei die Selektionsproblematik nachweisbar.

Vergleich der 95 %-Konfidenzintervalle. Der Vergleich der Konfidenzintervalle zeigt im Vergleich zum vorangegangenen Abschnitt nicht so große Intervallbreiten der Bildungsrenditen wie bei der separierten Betrachtung. Die genauen Intervalle pro Bundesland sind in Tabelle 6.25 sowie in Abbildung 6.12 veranschaulicht.

Bundesland	Bildungsrendite	Konfidenzintervall	Bundesland	Bildungsrendite	Konfidenzintervall
BW	0,076***	0,065–0,086	NI	0,070***	0,056–0,084
BY	0,074***	0,064–0,085	NW	0,079***	0,071–0,088
BE	0,091***	0,069–0,113	RP / SL	0,080***	0,063–0,096
BB	0,086***	0,065–0,107	SN	0,082***	0,069–0,098
HB	0,081***	0,044–0,118	ST	0,103***	0,084–0,122
HH	0,063***	0,031–0,091	SH	0,069***	0,044–0,094
HE	0,077***	0,063–0,091	TH	0,105***	0,084–0,126
MV	0,135***	0,106–0,164			

Tabelle 6.25: 95 %-Konfidenzintervalle aus der Heckman-Regression mit Interaktionstermen für die Bundesländer; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; Quelle: DIW, eigene Berechnungen.

Das Intervall von Mecklenburg-Vorpommern überschneidet sich, mit Ausnahme von Bremen, mit keinem weiteren westdeutschen Bundesland. Im Vergleich mit den ostdeutschen Bundesländern ist auch in diesem Fall die Überschneidungsfreiheit mit Sachsen gegeben. Alle weiteren Intervalle überschneiden sich. Demnach kann ein weiteres Mal die Unterschiedlichkeit von Mecklenburg-Vorpommern vermutet werden. Belegt werden kann dies auf Basis des Wald-Tests, der im nächsten Unterabschnitt dargestellt ist.

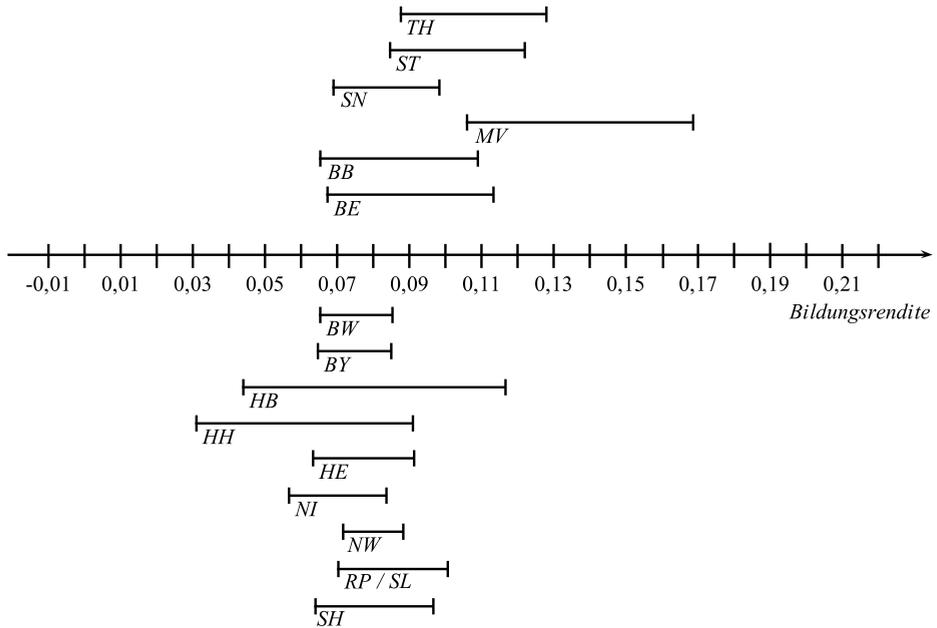


Abbildung 6.12: 95 %-Konfidenzintervalle für die mit dem Heckman-Verfahren ermittelten Bildungsrenditen pro Bundesland (IAT); Quelle: DIW, eigene Darstellung.

Vergleich über den Wald-Test. In Tabelle 6.26 werden die P-Werte aus dem Wald-Test der Bildungsrenditen gegeneinander angegeben. Dabei werden jene Koeffizienten kursiv gekennzeichnet für die gilt, dass die Gleichheitshypothese mit einer Irrtumswahrscheinlichkeit von 5 % abzulehnen ist.

Es zeigt sich beim Vergleich der Ergebnisse, dass sich die meisten Bundesländer im Hinblick auf die Bildungsrendite nur wenig unterscheiden und die Hypothese selten abzulehnen ist. Auch an dieser Stelle ist festzuhalten, dass sich die Bildungsrendite von Mecklenburg-Vorpommern – als einzigem Bundesland – signifikant von fast allen anderen unterscheidet. Lediglich für Thüringen und Sachsen-Anhalt kann kein signifikanter Unterschied geschätzt werden. Vergleicht man die Bildungsrenditen der letztgenannten mit den anderen Bundesländern, so ist die Verschiedenartigkeit in der Rendite nicht mehr ganz so deutlich nachzuweisen. Für Sachsen-Anhalt können im Vergleich zu allen westdeutschen Bundesländern, mit Ausnahme von Rheinland-Pfalz, Unterschiede gezeigt werden. Für den Vergleich mit Thüringen ist die Gleichheitshypothese für Hamburg, Hessen, Niedersachsen, Nordrhein-Westfalen, Sachsen und Schleswig-Holstein nachweisbar. Dementsprechend ist

6.3 Regionale Bildungsrenditen unter Berücksichtigung der Erwerbslosen

Wald-Test	BW	BY	BE	BB	HB	HH	HE	MV	NI	NW	RP/SL	SN	ST	SH	TH
BW	–														
BY	0,85	–													
BE	0,22	0,18	–												
BB	0,40	0,34	0,75	–											
HB	0,79	0,74	0,66	0,82	–										
HH	0,36	0,40	0,11	0,18	0,41	–									
HE	0,87	0,75	0,30	0,50	0,85	0,33	–								
MV	0,00	0,00	0,02	0,01	0,03	0,00	0,00	–							
NI	0,52	0,62	0,12	0,22	0,59	0,58	0,48	0,00	–						
NW	0,59	0,45	0,34	0,58	0,94	0,24	0,78	0,00	0,26	–					
RP / SL	0,69	0,59	0,42	0,65	0,94	0,28	0,82	0,00	0,38	0,98	–				
SN	0,49	0,40	0,53	0,76	0,95	0,21	0,62	0,01	0,25	0,75	0,82	–			
ST	0,01	0,01	0,41	0,24	0,30	0,02	0,03	0,07	0,01	0,03	0,07	0,10	–		
SH	0,62	0,69	0,19	0,31	0,60	0,68	0,57	0,00	0,93	0,43	0,48	0,37	0,03	–	
TH	0,02	0,01	0,36	0,21	0,27	0,02	0,03	0,10	0,01	0,03	0,06	0,09	0,89	0,03	–

Tabelle 6.26: Ergebnisse des Wald-Tests auf Unterschiedlichkeit der Bildungsrenditen aus der Heckman-Regression (IAT); Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; die kursiv geschriebenen Zahlen lassen die Ablehnung der Gleichheitshypothese zu. Quelle: DIW, eigene Berechnungen.

festzustellen, dass im Besonderen die Bildungsrendite von Mecklenburg-Vorpommern von den meisten anderen deutlich abgrenzbar ist. Aber auch für Sachsen-Anhalt und Thüringen kann im Vergleich zu fast allen anderen deutschen Bundesländern eine signifikant höhere Bildungsrendite aufgezeigt werden.

6.3.3 Zusammenfassung

In diesem Abschnitt werden bei der Schätzung regionaler Bildungsrenditen alle Erwerbspersonen berücksichtigt. Hierbei können die zuvor verwendeten Verfahren nicht zur Anwendung kommen. Um zusätzlich die nicht Erwerbspersonen zu berücksichtigen, wird das Heckman-Verfahren verwendet. Dieses wird sowohl auf einen regional separierten Datensatz als auch auf einen gesamtdeutschen Datensatz mit der regionalen Berücksichtigung durch Interaktionsterme angewandt. Für beide Verfahren ergibt sich, dass im Vergleich Mecklenburg-Vorpommern mit Abstand die höchsten Bildungsrenditen aufweist. Dabei zeigt sich die Bildungsrendite für dieses Bundesland über den separierten Datensatz in Höhe von 13,4 %, aus dem gesamtdeutschen Datensatz in Höhe von 14,5 %. Unabhängig von der Methode oder der Datenauswahl weist Mecklenburg-Vorpommern stets die höchsten Renditen auf. Im Vergleich zu den meisten anderen Bundesländern kann hier die höhere Rendite auch signifikant bestätigt werden. Zusätzlich zeigen sich auch für Sachsen-Anhalt und Thüringen hohe Renditen. Auch sie sind im Vergleich mit einigen Bundesländern signifikant verschieden nachweisbar. Demnach ist festzustellen, dass sich Mecklenburg-Vorpommern im Besonderen, aber auch Sachsen-Anhalt und Thüringen durch höhere Renditen auszeichnen. Alle anderen Bundesländer zeigen im direkten Vergleich keine signifikanten Unterschiede. Besonders die westdeutschen Bundesländer im direkten Vergleich zeigen in Bezug auf die Bildungsrendite deutliche Gemeinsamkeiten.

6.4 Vergleich der Schätzergebnisse und Zusammenfassung

Das zusammenfassende Kapitel stellt zunächst die verschiedenen Methoden und Modelle zusammenfassend dar und diskutiert im nächsten Schritt die Gemeinsamkeiten und Unterschiede der ermittelten Bildungsrenditen.

6.4.1 Vergleich der verwendeten Methoden

Basierend auf den Voraussetzungen des OLS-Verfahrens wird eine Modellspezifikation entwickelt, die die bestmögliche Schätzung verspricht. Nach durchgeführter Prüfung besagter Voraussetzungen zeigt sich ein Endogenitätsproblem für die Bildungsdauer, was zur Verzerrung des Koeffizienten der Bildungsrendite führen kann. Um dem zu begegnen wird das 2SLS-Verfahren basierend auf den notwendigen Annahmen diskutiert und eine zusätzliche Modellspezifikation für die Erklärung der Bildungsdauer entwickelt. Die Prüfung der Voraussetzungen zeigt, dass dieses Verfahren besser zur Schätzung der Bildungsrendite geeignet ist. Jedoch muss vom Vorliegen einer Korrelation zwischen den Fehlertermen der beiden Regressionen ausgegangen werden. Dies widerspricht einer Annahme zur Durchführung dieses sequentiellen 2SLS-Verfahrens. Daher wird auf das 3SLS-Verfahren verwiesen, welches eine simultane Schätzung beider Regressionsgleichungen vornimmt. Nun kann der reine Effekt von einem zusätzlichen Bildungsjahr auf das Einkommen näherungsweise geschätzt werden, ohne dass individuelle Heterogenitäten bezüglich Fähigkeit und Motivation der Erwerbstätigen beeinflussend wirken. Unberücksichtigt bleiben jedoch die Erwerbslosen. Es ist anzunehmen, dass sich die Gruppen der Erwerbstätigen und der Erwerbslosen selbst selektieren und voneinander unterscheidbar sind. Das Ergebnis aus dem 3SLS-Verfahren kann demnach lediglich für die aktiv am Arbeitsmarkt Beschäftigten interpretiert werden. Um hingegen alle Erwerbspersonen mit zu berücksichtigen, wird das Heckman-Verfahren verwendet. Hierbei wird in einer ersten Regression das inverse Mills-Ratio berechnet, welches für die unterschiedlichen Eigenschaften zwischen Erwerbstätigen und Erwerbslosen in der nachfolgenden Hauptregression kontrollierend hinzugenommen wird. Eine Rendite direkt für die Gruppe der Erwerbslosen kann aufgrund der mangelnden Beobachtbarkeit des Reservationslohns nicht berechnet werden. Durch die Korrektur ist die berechnete Bildungsrendite aus dem Heckman-Verfahren näherungsweise für alle Erwerbspersonen zu interpretieren.

Die absoluten Ergebnisse für die Renditen sind aufgrund der verschiedenen Methoden mit jeweilig diskutierten Vor- und Nachteilen nur schwer miteinander vergleichbar, zumal die Erweiterung des Datensatzes zu einer unterschiedlichen Grundgesamtheit führt. Was hingegen verglichen werden kann, sind die pro Methode ermittelten Relationen der Bundesländer zueinander.

6.4.2 Vergleich der geschätzten Bildungsrenditen

Über das OLS-Verfahren ergibt sich eine gesamtdeutsche Bildungsrendite in Höhe von 8,5 %. Mit dem 3SLS-Verfahren²⁰⁶ ist das Ergebnis höher. Hier ergibt sich eine Rendite in Höhe von 10,2 %. Die Schätzung mit dem Heckman-Verfahren zeigt eine mittlere jährliche Verzinsung von 7,7 %.

Im nächsten Schritt werden Bildungsrenditen nach einzelnen Merkmalen auf Basis des 3SLS-Verfahrens²⁰⁷ berechnet. Hierbei werden im Besonderen verschiedene Kohorten, Arbeitnehmer im öffentlichen Dienst und in der Privatwirtschaft sowie Voll- und Teilzeitbeschäftigte untersucht. Es zeigt sich, dass die Bildungsrendite mit steigender Kohorte zunimmt. Für die Gruppe der 30–40-Jährigen kann eine mittlere jährliche Verzinsung von 10 % berechnet werden, wogegen die älteste Gruppe mit einer durchschnittlichen Rendite in Höhe von 13,1 % rechnen darf. Die mittlere Kohorte erzielt durchschnittlich eine Rendite in Höhe von 11,2 %. Der Vergleich der Konfidenzintervalle ergibt jedoch, dass sich alle Intervalle mit einer Irrtumswahrscheinlichkeit von 5 % überschneiden und demnach nicht von tatsächlich vorhandenen Unterschieden ausgegangen werden kann. Ein Mitarbeiter des öffentlichen Dienstes verdient im Mittel 8,7 % pro zusätzlichem Bildungsjahr mehr. Die Privatwirtschaft entlohnt hingegen ein zusätzliches Jahr an allgemeinem Humankapital mit 10,5 %, wobei auch beim Vergleich dieser 95 %-Konfidenzintervalle die Überschneidungsfreiheit als nicht gegeben angesehen werden kann. Ein Arbeitnehmer der weniger als 30 h vertraglich vereinbart hat, wird mit durchschnittlich 11,9 % Mehreinkommen pro zusätzlichem Bildungsjahr entlohnt. Dagegen haben die Vollzeitkräfte lediglich eine Bildungsrendite von 9,9 %. Auch in diesem Fall ist die Überschneidungsfreiheit der Konfidenzintervalle nicht gewährleistet. Demnach ist festzustellen, dass durchaus Unterschiede je nach Separierung berechnet werden können, diese aber in keinem Fall auf Basis der 95 %-Konfidenzintervalle als annähernd signifikant verschieden zu bezeichnen sind.

Im nächsten Schritt ist eine regionale Analyse durchgeführt worden, die zunächst die Ost-West-Regionen²⁰⁸ mit dem 3SLS-Verfahren untersucht und im Weiteren auf die Bundesländerebene ausgeweitet wird. Die Schätzung der ostdeutschen Bildungsrendite ergibt eine mittlere Verzinsung von 14,8 %, die westdeutsche hingegen liegt mit 10,1 % niedriger. Der Vergleich der 95 %-Konfidenzintervalle zeigt eine deutliche Überschneidungsfreiheit der

²⁰⁶Die Ergebnisse bei Verwendung des 2SLS- oder 3SLS-Verfahrens unterscheiden sich nicht wesentlich.

²⁰⁷Das Heckman-Verfahren wird für diese Merkmale nicht angewendet, da die Unterschiedlichkeit nicht gegeben ist und eine nochmalige Schätzung keinen großartigen Erkenntnisgewinn verspricht.

²⁰⁸Berlin zählt komplett zur Ostregion.

Intervallspannen. Demnach können die Höhen der Bildungsrenditen in Ost- und Westdeutschland als verschieden angesehen werden.

Bundesland	3SLS-Verfahren (sep. Datensatz)	OLS-Verfahren (sep. Datensatz)	OLS-Verfahren (ges. Datensatz (IAT))	Heckman-Verfahren (sep. Datensatz)	Heckman-Verfahren (ges. Datensatz (IAT))
BW	0,088***	0,080***	0,082***	0,076***	0,076***
BY	0,087***	0,084***	0,081***	0,081***	0,074***
BE	0,106***	0,100***	0,102***	0,085***	0,091***
BB	0,156***	0,088***	0,090***	0,076***	0,086***
HB	0,124***	0,093***	0,090***	–	0,081***
HH	0,042*	0,071***	0,067***	–	0,063***
HE	0,107***	0,079***	0,082***	0,076***	0,077***
MV	0,172***	0,142***	0,142***	0,126***	0,135***
NI	0,093***	0,078***	0,078***	0,076***	0,070***
NW	0,106***	0,089***	0,089***	0,080***	0,079***
RP / SL	0,092***	0,085***	0,085***	0,080***	0,080***
SN	0,137***	0,084***	0,084***	0,073***	0,082***
ST	0,138***	0,107***	0,107***	0,083***	0,103***
SH	0,079***	0,078***	0,078***	0,071***	0,069***
TH	0,113***	0,110***	0,110***	0,092***	0,105***

Tabelle 6.27: Regionale Bildungsrenditen im Vergleich; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; Quelle: DIW, eigene Berechnungen.

Die Erweiterung auf Bundesländerebene mit diesem Verfahren ist in Tabelle 6.27 in Spalte 2 gezeigt. Die nächste Spalte zeigt die Ergebnisse für die regionalen Bildungsrenditen des separierten OLS-Verfahrens. Dieses wurde als Vergleichsinstrument berechnet, da für die gesamtdeutsche Regression mit regionaler Variation über Interaktionsterme das 3SLS-Verfahren nicht anwendbar ist. Diese Ergebnisse auf Basis der Interaktionsterme sind in Spalte 3 zusammengefasst. Die Spalten 4 und 5 zeigen die regionalen Bildungsrenditen für Deutschland auf Basis des Heckman-Verfahrens. Spalte 4 zeigt dabei die Ergebnisse auf Basis von separierten Regressionen pro Bundesland. Spalte 5 stellt die regionalen Bildungsrenditen dar, die sich aus der Variationen über Interaktionsterme ergeben.

In Abbildung 6.13 werden die geschätzten Renditen vergleichend dargestellt. Tendenziell kann gezeigt werden, dass über die Heckman-Regressionen niedrigere, über die 3SLS-Regressionen höhere Renditen geschätzt werden als durch das OLS-Verfahren. Weiterhin

6 Regionale Bildungsrenditen

erweisen sich die Ergebnisse pro Bundesland als robust. Zwar ist tendenziell ein gewisser Spielraum in der Höhe der regionalen Renditen zu erkennen, vergleicht man jedoch die Regionen zueinander, so ergibt sich unabhängig von der Methode ein ähnliches Bild in der Rangfolge der Bundesländer.

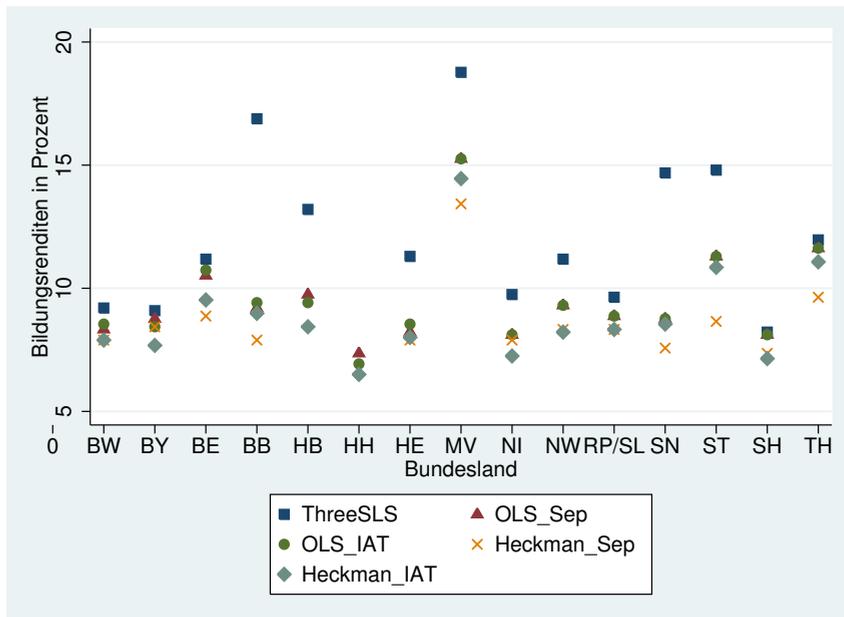


Abbildung 6.13: Bildungsrenditen im Vergleich; Quelle: DIW, eigene Berechnungen.

Es wird für Mecklenburg-Vorpommern stets die höchste Rendite, für Hamburg²⁰⁹ stets die niedrigste Rendite ermittelt. Die Ergebnisse variieren nur mäßig im Hinblick auf die Verwendung eines regional separierten bzw. gesamtdeutschen Datensatzes, unabhängig davon, ob die Schätzung mit dem OLS- oder Heckman-Verfahren durchgeführt wurde. Aufgrund der vielfältigen Probleme bei der generellen Schätzung von Bildungsrenditen, durch zum Beispiel nicht gegebene Quantifizierbarkeit von Motivation und Fähigkeit, sind diese Renditen als eine Tendenz zu verstehen.

²⁰⁹Die Bildungsrendite aus der 3SLS-Methode konnte für Hamburg nicht als signifikant von null verschieden bestätigt werden und ist damit nicht in der Grafik berücksichtigt. Zusätzlich sind die Beobachtungszahlen der Heckman-Methode mit regional separiertem Datensatz zu gering, um den Korrekturfaktor zu berechnen. Die Analyse wurde mit dieser Methode bei Separierung für Hamburg und Bremen nicht durchgeführt. Demnach ergeben sich für Hamburg nur drei geschätzte Renditen.

Allgemein ist festzustellen, dass auf Grundlage der Stichprobe die mittlere Bildungsrendite pro Bundesland im Intervall zwischen ca. 6,5 % und 18,8 % schwankt. Tendenziell ist die Rendite in den neuen Bundesländern höher als in den alten. Eine signifikant höhere Rendite im Vergleich zu den meisten Bundesländern ist für Mecklenburg-Vorpommern zu berechnen. Auch Sachsen-Anhalt und Thüringen tendieren zu einer höheren Rendite. Alle anderen ermittelten Bildungsrenditen pro Bundesland können als näherungsweise homogen zueinander verstanden werden.

6.4.3 Kritik

Die ermittelten Bildungsrenditen ergeben sich aus einem angenommenen semilogarithmischen Zusammenhang zwischen der Bildungsdauer und dem Einkommen. Sie unterstellen dabei, dass es für ein zusätzliches Bildungsjahr stets den gleichen Zuwachs an Einkommen gibt. Diese Annahme der Linearität über die Bildungsjahre ist eine notwendige Voraussetzung zur Schätzung der Renditen. Ob dieser Zusammenhang tatsächlich linear oder über einen anderen Verlauf besser abgebildet werden kann, ist jedoch nicht auszuschließen.

In diesem Kapitel werden die Rangfolgen der aus verschiedenen Methoden errechneten Bildungsrenditen gezeigt. Es ergibt sich jedoch, dass die absoluten Höhen der Renditen schon zwischen dem 3SLS- und OLS-Verfahren nicht vergleichbar sind, da von einem verzerrten Schätzer bei letzterem auszugehen ist. Weiterhin werden die Koeffizienten aus dem Heckman-Verfahren in den Vergleich mit einbezogen. Diese basieren jedoch auf einem um die Erwerbslosen erweiterten Datensatz und demnach einer anderen Grundgesamtheit. Doch unabhängig von Methode und Datensatz war es stets das Ziel, die Verzinsung eines zusätzlichen Bildungsjahres auf das Einkommen zu schätzen. Es kann angenommen werden, dass sich die zum Teil unberücksichtigten Einflüsse pro Methode in den Regionen ähnlich auswirken und demnach pro Verfahren eine ungefähre Rangfolge aufzustellen ist. Dabei spielen kleine Abweichungen in der Höhe der Bildungsrendite keine Rolle, aber die Tendenz für höhere oder niedrigere Renditen ist doch erkennbar. Die einzelnen Rangfolgen können nun mit Vorsicht verglichen werden. Da sich für alle Regressionen herausbildet, dass Mecklenburg-Vorpommern eine deutlich höhere Rendite hat als die der anderen Bundesländer, kann die Repräsentativität dieser Aussagen auch für die betrachtete Grundgesamtheit tendenziell vermutet werden.

Alle Ergebnisse basieren auf einer Stichprobe und wurden mit nahezu demselben Datensatz berechnet. Es kann nicht mit letzter Sicherheit ausgeschlossen werden, dass die sich

ergebenden durchschnittlichen Werte tatsächlich auf die Grundgesamtheit übertragbar sind. Im Vergleich zu weiteren empirischen Studien zeigen sich jedoch vergleichbare Ergebnisse, sodass es keinen Grund gibt, von einer grundlegenden Abweichung dieser Stichprobe auszugehen. Jedoch basieren die Vergleichsstudien ebenfalls alle auf stichprobenbasierten Daten, sodass auch für diese in letzter Konsequenz die garantierte Übertragbarkeit auf die Grundgesamtheit nicht gewährleistet ist. Tendenziell zeigen die Ergebnisse jedoch, dass es keinen Grund gibt, an der Repräsentativität der verwendeten Stichprobe zu zweifeln.

6.5 Regionale Unterschiede in Deutschland

In Kapitel 5 wurden mit unterschiedlichen Methoden und Modellspezifikationen Bildungsrenditen geschätzt. Mit Blick auf ausgewählte Merkmale konnten keine signifikant verschiedenen Renditen in Deutschland ermittelt werden. Bei der Schätzung regionaler Renditen in Kapitel 6 ergab sich zum einen eine deutliche Verschiedenartigkeit in den Bildungsrenditen zwischen Ost- und Westdeutschland. Zum anderen konnte diese Verschiedenartigkeit bei der weiteren Separierung nach Bundesländern im Besonderen für Mecklenburg-Vorpommern festgestellt werden. Für dieses Bundesland ergab sich – unabhängig vom Verfahren – eine deutlich höhere Rendite. Tendenziell höhere Renditen versprochen ebenso Sachsen-Anhalt und Thüringen. Diese Unterschiede sind zwar nicht für jedes Verfahren signifikant zu bestätigen, zeigen jedoch eine starke Tendenz. Alle anderen Renditen sind nicht voneinander statistisch abgrenzbar. Allerdings kann für alle ostdeutschen Bundesländer separiert eine tendenziell höhere Bildungsrendite geschätzt werden als für Westdeutschland. Im Vergleich der westdeutschen Bundesländer untereinander können die Renditen als annähernd gleich bezeichnet werden.

In diesem Abschnitt wird die mögliche Bedeutung der Unterschiede in den regionalen Bildungsrenditen diskutiert. Für die Vergleiche werden die Bildungsrenditen aus dem Heckman-Verfahren mit Interaktionstermen verwendet. Da mit allen Verfahren die Rangfolge zwischen den Bundesländern ähnlich ist, ist es letztlich nicht entscheidend aus welcher Methode die errechneten Renditen in dieser Analyse verwendet werden. Der Vorteil dieses Verfahrens ist jedoch, dass für alle Bundesländer eine signifikante Rendite geschätzt wurde und demnach die Diskussion mit Berücksichtigung aller Bundesländer erfolgen kann.

6.5.1 Bedeutung der regionalen Bildungsrendite

Die Bildungsrendite an sich gibt die mittlere Steigerung des Einkommens durch ein zusätzliches Bildungsjahr an. Ist die Rendite²¹⁰ in Ostdeutschland höher als in Westdeutschland, so bedeutet dies, dass ein Jahr zusätzliche Generierung von Humankapital eine größere Einkommenssteigerung im Osten erwirkt als im Westen. Diese bedeutet jedoch nicht, dass das absolute Einkommen in dieser Region höher ist. Im Gegenteil: Eine höhere Steigerung ist auf ein geringeres Anfangsniveau zurückzuführen. Chiswick (1968), S. 498, formuliert, dass das „*low level of income [...] of less developed regions increase the rate of return.*“ Demnach werden bei dem Vergleich der Bildungsrenditen nicht die tatsächlichen Einkommen miteinander verglichen, sondern lediglich die Bewertungen des zusätzlichen allgemeinen Humankapitals.

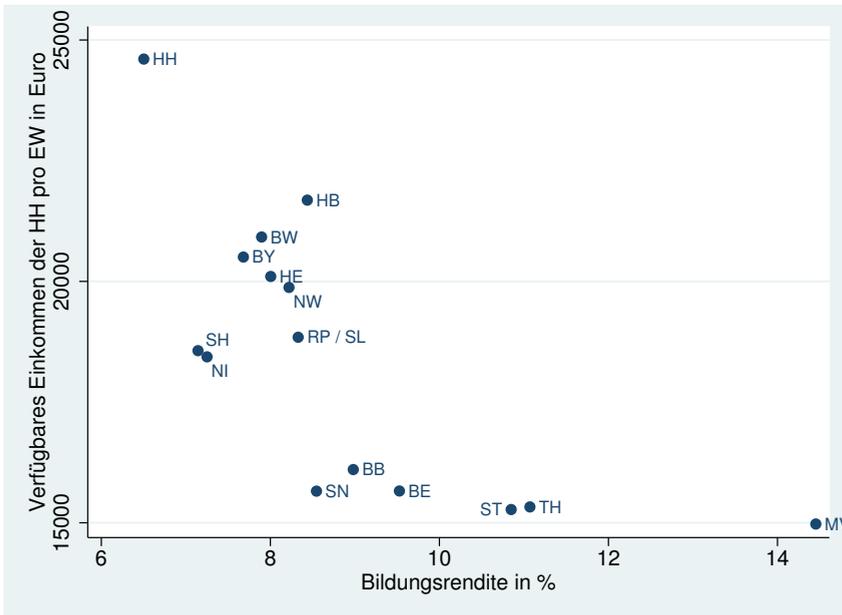


Abbildung 6.14: Regionales verfügbares Einkommen der Haushalte pro Einwohner und Bildungsrenditen; Quelle: Statistisches Bundesamt, DIW, eigene Berechnungen.

²¹⁰Eine Prüfung der Zusammenhänge erfolgt zusätzlich über den Korrelationskoeffizienten Kendalls-Tau, dargestellt im Anhang A.28.

In Abbildung 6.14 wird diese Relation auf Bundesländerebene dargestellt. Hierbei wird das allgemein verfügbare Einkommen der Haushalte pro Einwohner²¹¹ der ermittelten Bildungsrendite gegenübergestellt. Es liegt zwischen knapp 15.000 € in Mecklenburg-Vorpommern und knapp 25.000 € in Hamburg. Die vermutete Beziehung zwischen dem Einkommensniveau und der Bildungsrendite kann bestätigt werden. So ist eine Ost-West-Trennung zu erkennen. Mecklenburg-Vorpommern, Sachsen-Anhalt und Thüringen haben die geringsten verfügbaren Einkommen und sehr hohe Bildungsrenditen. Hamburg hat mit Abstand das höchste Einkommensniveau auf Bundesländerebene und weist die geringste Bildungsrendite auf. Wenn jedoch die Bildungsrenditen bei geringem Einkommenslevel hoch sind, ergibt sich daraus, dass die Spanne der gezahlten Einkommen groß ist und demzufolge die Abweichungen dieser von den mittleren Einkommen ebenfalls hoch sein müssen. Kuznet (1955), S. 5, schreibt diesbezüglich, dass eine *„reduction in the inequality of the percentage shares was accompanied by significant rises in real income per capita“*. Zur Prüfung dieses Zusammenhangs wird die mittlere quadratische Abweichung vom durchschnittlichen Bruttostundenlohn, d. h. die Standardabweichung, den errechneten Bildungsrenditen gegenübergestellt. Hierbei ist die Verwendung des Variationskoeffizienten erforderlich, da ansonsten die unterschiedlichen Höhen der Einkommen in den Bundesländern berücksichtigt wären. Durch den Variationskoeffizienten wird die relative Abweichung angegeben, die wiederum als Prozentzahl zwischen den Bundesländern vergleichbar ist.

In Abbildung 6.15 wird der Variationskoeffizient der Einkommen in Relation zu den regionalen Bildungsrenditen dargestellt. Hierbei ist keine klare Ost-West-Teilung erkennbar. Tendenziell ähnliche Einkommen haben Schleswig-Holstein, Bremen und Baden-Württemberg mit 40 %–43 % Abweichung. Eine sehr ungleiche Einkommensverteilung ist hingegen für Mecklenburg-Vorpommern mit über 70 % mittlerer quadratischer Abweichung vom durchschnittlichen Einkommen zu verzeichnen. Auch Sachsen-Anhalt zeigt mit etwa 60 % den zweithöchsten Variationskoeffizienten.

Es ist feststellbar, dass in Mecklenburg-Vorpommern, dem Land mit der mit Abstand höchsten Bildungsrendite, die generellen Einkommen geringer sind als in anderen Bundesländern. Dagegen jedoch fallen auch die relativen Abweichungen zum Einkommen deutlich höher aus. Der Grund hierfür liegt in der größeren Mobilität Hochqualifizierter.²¹² Demnach wird ein Hochqualifizierter die Arbeitsplatzangebote eher überregional vergleichen. Daraus ergibt sich, dass sich die Einkommen auch in einkommensschwachen

²¹¹Die Daten stammen vom statistischen Bundesamt für das Jahr 2008.

²¹²Vgl. u. a. Chiswick (1968).

Regionen über die Bundesländergrenzen hinaus annähern müssen, um eine Abwanderung dieser Gruppe zu verhindern. Eine absolute Angleichung der Einkommen bedeutet dies jedoch nicht zwangsläufig.

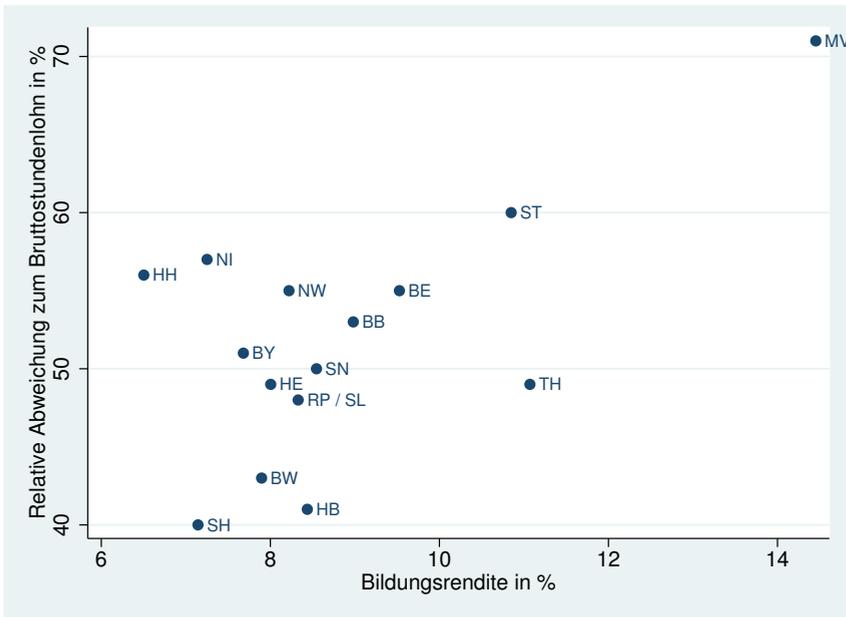


Abbildung 6.15: Regionale relative Standardabweichung des Bruttostundenlohns und Bildungsrenditen; Quelle: DIW, eigene Berechnungen.

Werden im nächsten Schritt nun die Wanderungsbewegungen von Hoch- und Geringqualifizierten betrachtet, so ergibt sich folgendes Bild. In den Abbildungen 6.16 werden die Nettomigrationsquoten Hoch- und Geringqualifizierter²¹³ gegenübergestellt. Dabei sind die Wanderungssalden der Jahre 2000–2006 berücksichtigt. In Abbildung 6.16.a wird der Anteil der Wanderungen Hochqualifizierter gezeigt. Hierbei ist zu erkennen, dass diese tendenziell nach Berlin, Schleswig-Holstein und Bayern wandern – die Nettomigrationsquote dieser Gruppe in Berlin liegt bei über 13 ‰. Abwanderungen Hochqualifizierter erfolgen aus den Ländern Brandenburg, Sachsen-Anhalt und Thüringen. Alle weiteren Länder haben mit einem Zu- und Abwanderungssaldo von ± 5 ‰ eine recht konstante Zahl an Hochqualifizierten. Zu dieser Gruppe zählt auch Mecklenburg-Vorpommern. Die

²¹³Die Daten stammen aus Granato und Niebuhr (2009). Sie verwenden IAB-Daten. Dabei werden ausschließlich die Inlandswanderungen von Vollzeitbeschäftigten berücksichtigt. Hochqualifizierte sind Personen mit einem tertiären Abschluss, Geringqualifizierte sind all jene, die keine abgeschlossene Berufsausbildung haben. Das Wanderungssaldo wird dabei berechnet aus $\text{gew. Wanderungssaldo} = \frac{\text{Einwanderer} - \text{Auswanderer}}{\text{Beschäftigte am Arbeitssort}} \cdot 1000$.

6 Regionale Bildungsrenditen

Nettomigrationsquote der Hochqualifizierten ist hier mit -4 ‰ vergleichbar mit der von Bremen. Die Hochqualifizierten haben demzufolge durch die höhere Rendite eher einen Anreiz, in dieser einkommensschwachen Region zu bleiben.

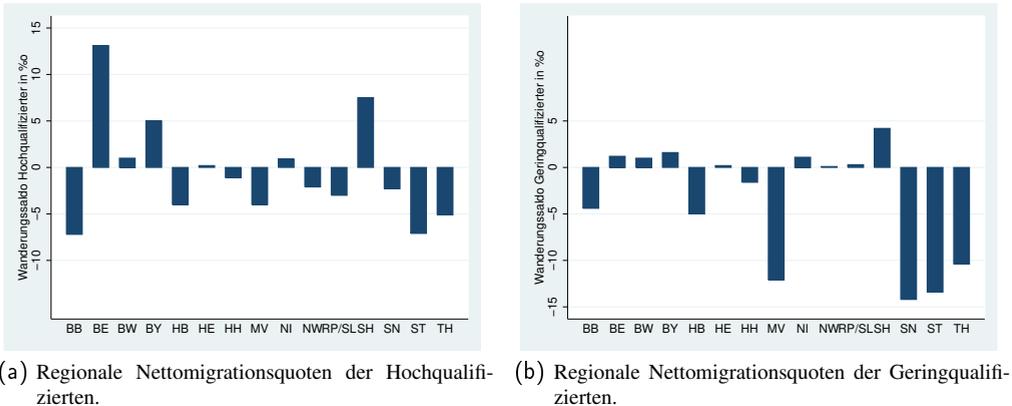


Abbildung 6.16: Wanderungssalden der Bundesländer zwischen 2000 und 2006; Quelle: IAB, eigene Darstellung.

Betrachtet man hingegen die Nettomigrationsquoten der Geringqualifizierten, so ist wieder eine deutliche Ost-West-Trennung auszumachen. Aus Sachsen, Sachsen-Anhalt, Mecklenburg-Vorpommern und Thüringen wandern deutlich mehr Geringqualifizierte ab als aus allen anderen Bundesländern.

Die Nettomigrationsquote zeigt für diese Bundesländer eine Abwanderungsquote zwischen knapp 10 ‰ und 15 ‰. Für Bremen und Brandenburg ist eine Nettomigration von ca. -5 ‰ zu erkennen. Alle weiteren Bundesländer haben nur eine sehr geringe Abwanderungsquote, oder aber eine leicht positive Nettomigrationsquote unter den Geringqualifizierten.

Der Grund für die hohe Zahl an Abwanderungen von Geringqualifizierten könnte durch ein hohes Arbeitslosigkeitsrisiko für die Gruppe gegeben sein. In einer strukturschwachen Region wie Mecklenburg-Vorpommern, besteht vermutlich ein hohes Arbeitsangebot von Geringqualifizierten. Ein Indiz dafür liefern die Arbeitslosenquoten. Daten für diese je nach Qualifikation und Region sind nicht erhältlich, daher wird in Abbildung 6.17 die allgemeine regionale Arbeitslosenquote in Relation zur Bildungsrendite dargestellt.

Eine geringe Arbeitslosenquote deutet auf eine gegebene Knappheit von Arbeitnehmern hin. Diese wird sich je nach Qualifikationsniveau unterscheiden. Konkret ist davon auszu-

gehen, dass die Arbeitslosenquote für Geringqualifizierte deutlich höher ist als für Hochqualifizierte. Dies bedeutet, dass die in Abbildung 6.17 dargestellten Anteile der Erwerbslosen an allen Erwerbstätigen zu einem Großteil von Geringqualifizierten bestimmt werden. Auch in dieser Abbildung ist wieder eine klare Ost-West-Teilung zu erkennen. Die höchsten Arbeitslosenquoten ergeben sich für Mecklenburg-Vorpommern, Sachsen-Anhalt und Berlin. Einzige Ausnahme bildet Bremen, welches sich mit einer Arbeitslosenquote von knapp 12 % deutlich von den anderen westdeutschen Bundesländern abhebt.

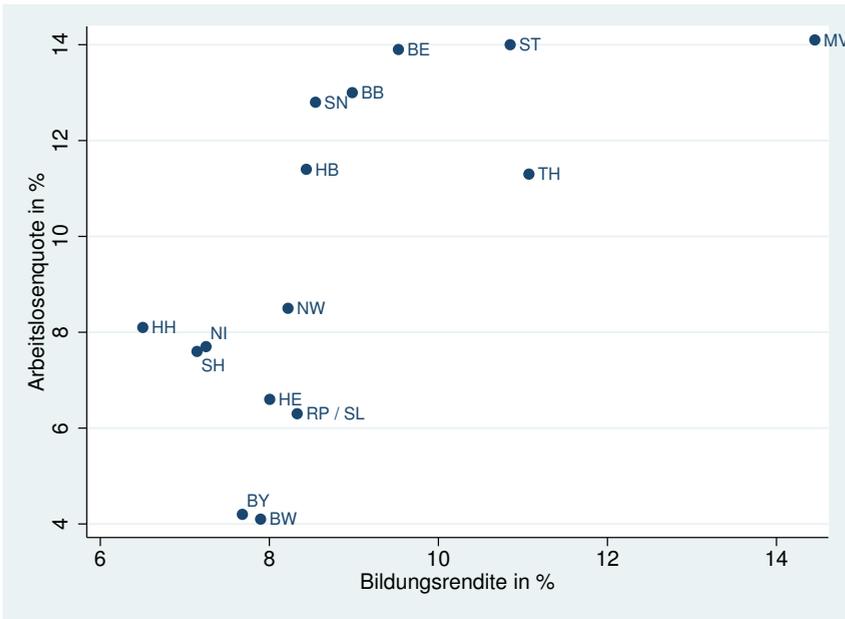


Abbildung 6.17: Regionale Arbeitslosenquoten und Bildungsrenditen; Quelle: Statistisches Bundesamt, DIW, eigene Berechnungen.

IAB-Aktuell (2011) zeigt Arbeitslosenquoten für die Ost- und West-Region separiert nach Qualifikationen für das Jahr 2009. So ist die Arbeitslosenquote für Personen ohne Berufsabschluss im Osten bei 34,0 %, im Westen bei 20,1 %. Personen mit einem Hoch- bzw. Fachhochschulabschluss sind im Osten lediglich mit einer Wahrscheinlichkeit von 4,2 %, im Westen mit 2,0 % arbeitslos. Auch die Zahlen belegen die Vermutung, dass sich die Wanderungsbewegung der Geringqualifizierten auf eine hohes Arbeitsangebot für diese Gruppe in den neuen Ländern zurückführen lässt.

6 Regionale Bildungsrenditen

Das generell geringe Einkommensniveau lässt sich nun auf das große Angebot an geringqualifizierten Arbeitskräften zurückführen. Für sie bleibt die Alternative, das geringe Einkommenslevel zu akzeptieren oder aber die Migration in andere Bundesländer. Die Akzeptanz des deutlich geringeren Einkommensniveaus ist nur möglich, wenn auch mit diesem die Lebenshaltungskosten gedeckt werden können. Daher ist davon auszugehen, dass eine hohe Bildungsrendite einhergeht mit einem geringen durchschnittlichen Preisniveau.

Daten des regionalen Preisniveaus sind auf Bundesländerebene nicht erhältlich. Daher werden zumeist Mieten oder Kaufpreise für Grund und Boden²¹⁴ als Ersatz herangezogen. In dieser Arbeit wird dabei auf den durchschnittlichen Kaufpreis pro Quadratmeter für alle Arten von Bauland²¹⁵ zurückgegriffen. Dabei werden in Abbildung 6.18 die Stadtstaaten aufgrund einer nicht gegebenen Vergleichbarkeit mit den Flächenländern herausgenommen und lediglich die mittleren Baulandpreise der Flächenländer dargestellt.

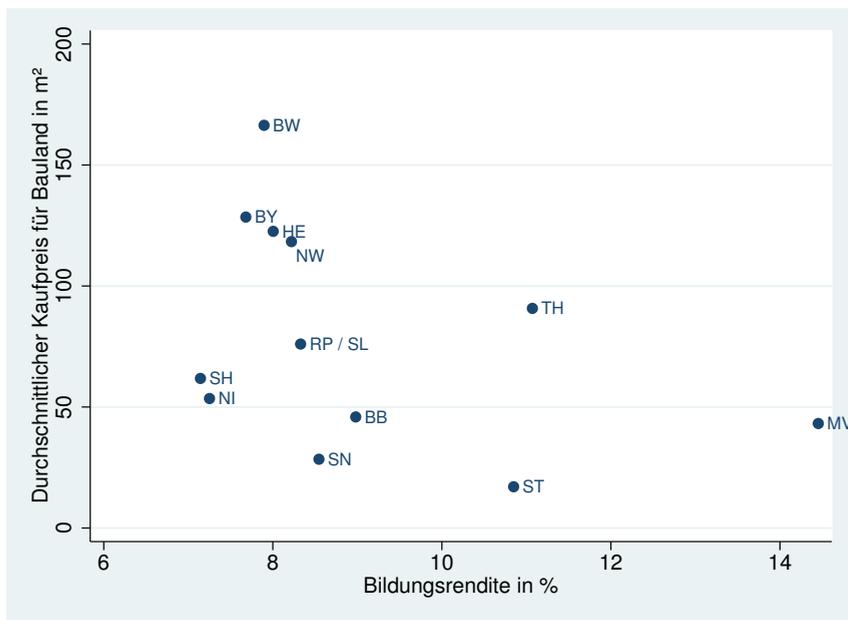


Abbildung 6.18: Regionale Baulandpreise und Bildungsrenditen; Quelle: Statistisches Bundesamt, DIW, eigene Berechnungen.

²¹⁴Eine Diskussion über deren Verwendung findet sich in Von der Lippe und Breuer (2010).

²¹⁵Die Daten stammen aus den Analysen des statistischen Bundesamtes für das Jahr 2008.

Die vermutete deutliche Abgrenzbarkeit zwischen den ost- und westdeutschen Ländern ist nicht zu bestätigen. Allerdings kann tendenziell bei einer geringen Bildungsrendite ein hoher Baulandpreis identifiziert werden. Für die meisten ostdeutschen lassen sich geringere Kaufpreise pro Quadratmeter aufzeigen als bei den westdeutschen Bundesländern. Die geringsten Baulandpreise ergeben sich für Sachsen-Anhalt, die höchsten für Baden-Württemberg.

Generell lässt sich aus den bivariaten Betrachtungen eine negative Beziehung zwischen der Bildungsrendite und dem verfügbaren Einkommen ableiten. Demzufolge ergibt sich auch eine größere Schwankung in den Einkommen der Bundesländer mit hoher Rendite. Im Gegenzug ist festzuhalten, dass es eine positive Korrelation zwischen der Arbeitslosenquote und der Bildungsrendite gibt. Dies deutet auf eine größere Knappheit von vorrangig Geringqualifizierten in den Bundesländern mit geringer Bildungsrendite hin. Dadurch ist das generell höhere Einkommensniveau in diesen Bundesländern eine logische Folge. Diese Bundesländer sind tendenziell attraktiv für Geringqualifizierte, wohingegen durch eine hohe Bildungsrendite die Hochqualifizierten tendenziell in den einkommensschwachen Bundesländern gehalten werden können. Dies lässt sich auch über den Vergleich der Wanderungssalden zeigen. Für die Hochqualifizierten gilt zudem, dass die Kaufkraft, trotz geringerem absoluten Einkommen, durch ein geringeres generelles Preisniveau in diesen Ländern nicht schlechter ist als in anderen Bundesländern.

6.5.2 Kritik

Mit den bivariaten Vergleichen ist es möglich, Zusammenhänge zwischen den regionalen Bildungsrenditen und einigen makroökonomischen Größen aufzuzeigen. So kann eine Relation zum verfügbaren Einkommen, den relativen Abweichungen dieser Einkommen, zu den Arbeitslosenquoten oder auch den Baulandpreisen dargestellt werden. Die intuitiven Annahmen können bestätigt werden. Problematisch jedoch ist, dass es sich bei dieser Analyse lediglich um einen bivariaten Vergleich von zwei Stichproben handelt. Eventuelle Scheinkorrelationen aufgrund von dritten unberücksichtigten Einflüssen können nicht ausgeschlossen werden.

Die zum Vergleich verwendeten Stichproben werden in jedem Fall als repräsentativ angenommen, sodass die Daten und Ergebnisse miteinander vergleichbar sind. Dies ist gerade mit Blick auf die verschiedenen Alterseingrenzungen eine restriktive Annahme. Jedoch ist anzunehmen, dass sich die Relationen auch in anderen Altersgrenzen nicht wesentlich

ändern. So ist auch für die Gruppe der über 60-Jährigen eine höhere Arbeitslosenquote der Gering- als der Hochqualifizierten anzunehmen.

Aus dem Vergleich der Stichproben können lediglich Tendenzen abgeleitet werden. Eine tatsächlich statistisch fundierte Analyse ist mit 15 Datenpunkten²¹⁶ nicht durchführbar. Der Einfluss auf diverse makroökonomische Größen ist vielfältig und kann durch eine rein bivariate Betrachtung nicht in Gänze abgebildet werden. Eine Ursache- Wirkungskette ist damit nicht aufzubauen. Das Ziel dieses Kapitels kann es daher nur sein, mögliche tendenzielle Zusammenhänge aufzuzeigen. Eine abschließende Analyse wird nicht angestrebt. Für eine fundierte Analyse wäre eine mehrperiodische Betrachtung zur Vergrößerung der Datenmenge notwendig.

Im Fall der Wanderungsbewegungen musste, mangels Verfügbarkeit, auf Daten von 2000–2006 zurückgegriffen werden. Die Auswertung und auch alle anderen Daten stammen aus dem Jahr 2008. Somit muss für erstgenannte die Annahme eines sich fortsetzenden Trends bis 2008 unterstellt werden.

Eine Variable für das regionale Preisniveau ist nicht erhältlich. Daher werden näherungsweise die durchschnittlichen Baulandpreise angenommen. Besonders mit Blick auf die Stadtstaaten ist das knappe Gut Boden nicht mit den Flächenländern vergleichbar. Demzufolge wurden sie in dieser Analyse unberücksichtigt gelassen. Zudem hängen die durchschnittlichen Baulandpreise von der jeweiligen Struktur des Landes ab. So ist der mittlere Kaufpreis in Nordrhein-Westfalen aufgrund einer verschiedenen Landesstruktur nicht mit der von beispielsweise Bayern zu vergleichen. Demzufolge kann diese Hilfsvariable lediglich als Indiz für die Vermutung des geringeren Preisniveaus in den ostdeutschen Bundesländern dienen.

6.5.3 Zusammenfassung

Zusammenfassend ist festzustellen, dass die regionalen Unterschiede in Höhe der Bildungsrendite mit einem generell geringeren Einkommensniveau einhergehen. Dieses ist nur aufgrund eines geringeren regionalen Preisniveaus – hier näherungsweise über mittlere Baulandpreise abgebildet – möglich. Die relativen Abweichungen der Einkommen bestätigen die Unterschiede in den Bundesländern. Sie zeigen, dass im Bundesland mit der höchsten Rendite die höchste relative Abweichung vom mittleren Bruttostundenlohn gegeben ist. Weiterhin ist festzustellen, dass in den Ländern mit hoher Bildungsrendite

²¹⁶Berechnet wurden jeweilig 15 Bildungsrenditen.

eine hohe Arbeitslosigkeit existiert, was auf ein hohes Arbeitsangebot für Geringqualifizierte hindeutet und die Akzeptanz eines geringeren Einkommens erhöht. Doch durch das hohe Angebot besteht die Notwendigkeit für eine überdurchschnittliche Abwanderung dieser Gruppe aus den ostdeutschen Bundesländern. Diese Wanderungsbewegung ist für die Hochqualifizierten nicht in dem Maße zu erkennen. Dies ergibt sich, da sie zum einen aufgrund einer anzunehmenden deutlich geringeren Arbeitslosenquote weniger die Notwendigkeit haben, auf der anderen Seite durch eine deutlich höhere monetäre Verzinsung von Humankapital motiviert werden, im Bundesland zu bleiben.

7 Schlussbetrachtung

In dieser Arbeit wurde zunächst die theoretische Fundierung von Bildungsrenditen diskutiert. Hierfür sind verschiedene Modelle von Jacob Mincer vorgestellt worden. Am Ende ergab sich eine technische Formulierung zur Erklärung des Einkommens durch die Bildungsdauer und die Berufserfahrung. Aufbauend auf dieses Modell sind diverse empirische Schätzungen bereits vorgenommen worden. Dabei schwanken die ermittelten Bildungsrenditen aus den meisten Studien zwischen 6,5 % und 10,2 %, wobei sich die Berechnungen je nach verwendetem Verfahren tendenziell leicht unterscheiden.

Als Grundlage für die Analyse wurde ein Datensatz vom Deutschen Institut für Wirtschaftsforschung verwendet. Diese Daten des SOEP sind in einer mehrperiodischen Erhebung seit 1984 auf Individuallevel erhoben worden. Dabei werden die gleichen Personen und Haushalte jährlich befragt. Der Grunddatensatz wurde reduziert auf die 30–60-Jährigen nicht selbstständigen Erwerbstätigen, die nicht im Ausbildungsprozess sind und zu allen relevanten Variablen Angaben gemacht haben. Die Repräsentativitätsprüfung ist mit dem Mikrozensus als Vergleichsdatsatz durchgeführt worden. Der Vergleich einiger für die Analyse bedeutender Variablen ergab, dass der Datensatz für Erwerbstätige in der ausgewählten Altersgruppe annähernd als repräsentativ angenommen werden konnte.

Aufbauend auf die theoretisch diskutierte Lohngleichung wurden mögliche Modellspezifikationen diskutiert. Das Ziel war es hierbei, den Einfluss von allgemeinem Humankapital zu ermitteln. Dieser Koeffizient wurde zunächst über das Mincer-Modell via OLS-Verfahren geschätzt und dann schrittweise um einige Variablen erweitert, um einen möglichst genauen Schätzer der Bildungsrendite zu ermitteln. Diese verschiedenen Regressionen wurden im nächsten Schritt mit Blick auf die Erfüllung der Voraussetzung für dieses Verfahren geprüft. Bei einigen dieser Voraussetzungen war die Erfüllung essentiell, um den besten linearen unverzerrten Schätzer zu ermitteln, andere konnten bei Nichterfüllung über einzelne Korrekturen behoben werden. Es ergab sich durch die mangelnde Quantifizierbarkeit des Einflusses individueller Fähigkeiten und Motivationen, dass der direkte Effekt der reinen Bildungsdauer auf das Einkommen nicht unverzerrt geschätzt

7 Schlussbetrachtung

werden kann. Zur Lösung dieses Problems wurde das 2SLS-Verfahren diskutiert. Hierbei wird in einem sequentiellen Verfahren eine individuelle Bildungsdauer geschätzt, bei der möglichst der Einfluss der Fähigkeiten unberücksichtigt bleibt. Durch dieses zweistufige Schätzverfahren ergab sich jedoch bei den verwendeten Modellspezifikationen eine Korrelation zwischen den beiden Fehlertermen. Um diesem zu begegnen, wurde das simultane 3SLS-Verfahren angewendet. Hierbei konnte nun näherungsweise der verzerrende Effekt herausgerechnet und die zuvor kritisierte Korrelation durch die Simultanität eliminiert werden. Aus diesem Verfahren konnte eine gesamtdeutsche Bildungsrendite in Höhe von 10,2 % pro zusätzlichem Bildungsjahr ermittelt werden. Mit diesem Verfahren sind nun zunächst Bildungsrenditen nach drei ausgewählten Merkmalen untersucht worden. Dabei wurden Unterschiede bedingt durch eine verschiedene Altersstruktur betrachtet. Die zweite Unterscheidung diskutierte die verschiedenen Einflüsse der Privatwirtschaft bzw. des öffentlichen Dienstes, und in der dritten Untersuchung fand eine Separierung nach Voll- und Teilzeiterwerbstätigen statt. Für alle Merkmalsunterscheidungen konnten über den Vergleich der 95 %-Konfidenzintervalle keine signifikant verschiedenen Bildungsrenditen geschätzt werden.

Die Hinzunahme der Erwerbslosen in einem vergrößerten Datensatz diente der Erweiterung der Aussagekraft. Hierfür wurde zunächst der Datensatz um ALG-I-Empfänger erweitert. Der Einbezug dieser Gruppe änderte sowohl bei gemeinschaftlicher als auch bei einer separierten Betrachtung nichts an der nationalen Bildungsrendite. Daher war eine weitere Verfolgung irrelevant. Um die ALG-II-Empfänger zusätzlich zu berücksichtigen, musste ein weiteres Verfahren angewendet werden. Mit dem Verfahren von James Heckman wurde nun in einer vorgelagerten Probit-Regression ein Korrekturfaktor berechnet, der die unterschiedlichen Charakteristika zwischen Erwerbstätigen und Erwerbslosen berücksichtigt. Unter Einbezug dieses Faktors wurde die Lohngleichung ein weiteres Mal geschätzt, sodass die verzerrenden Einflussfaktoren annähernd herauszurechnen waren. Mit dieser Methode konnte nun eine gesamtdeutsche Bildungsrendite in Höhe von 7,7 % geschätzt werden.

Im nächsten Abschnitt wurden regionale Unterschiede in Deutschland näher untersucht. Eine Ost-West-Separierung ergab eine deutliche Abweichung der 95 %-Konfidenzintervalle der separat ermittelten Bildungsrenditen. Es konnte eine deutlich höhere Rendite für die ostdeutschen als für die westdeutschen Bundesländer geschätzt werden. Bei weiterer Separierung auf Bundesländerebene zeigte sich diese Struktur nicht mehr so klar. Die Gegenüberstellung der Konfidenzintervalle ergab im Vergleich von Mecklenburg-Vorpommern

mit den westdeutschen Bundesländern zumeist eine komplette Überschneidungsfreiheit. Aufgrund der Separierung des Datensatzes und der Durchführung von Einzelregressionen konnte jedoch kein statistischer Test auf Gleichheit der Koeffizienten erfolgen. Das wurde in einem neuen Regressionsmodell realisiert. Damit wurden auf Basis eines gesamtdeutschen Datensatzes regionale Bildungsrenditen geschätzt. Dabei werden die Bundesländer über Interaktionsterme variiert. Dieses neue Verfahren ist mit der 3SLS-Methode nicht durchführbar, da sich 15 Zusatzregressionen ergäben, die bei begrenzter Anzahl von Instrumenten zur Erklärung der Bildungsdauer keine Identifikation des Zusatzmodells zulassen. Daher wird diese Schätzung mit dem OLS-Verfahren durchgeführt. Die absoluten Höhen der Bildungsrenditen änderten sich im Vergleich zu der separierten Betrachtung, aber die Reihenfolge zwischen den Bundesländern, mit der höchsten Rendite für Mecklenburg-Vorpommern, war auch mit dieser Methode nachweisbar. Der auf das gesamtdeutsche Modell anwendbare Wald-Test zur Überprüfung der Gleichheit der Koeffizienten bestätigte die via 3SLS-Verfahren aufgestellte Vermutung. Mit Ausnahme des bilateralen Vergleichs zu Sachsen-Anhalt und Thüringen konnte für alle anderen regionalen Vergleiche mit Mecklenburg-Vorpommern eine signifikant verschiedene Bildungsrendite berechnet werden. Der Vergleich der beiden erstgenannten Bundesländern mit den restlichen – außer Mecklenburg-Vorpommern – zeigte diese Ablehnbarkeit der Gleichheitshypothese nicht so klar, war aber noch zu großen Teilen gegeben. Demzufolge konnte aus dieser Analyse eine deutliche Unterscheidung der Bildungsrenditen von Mecklenburg-Vorpommern im Besondern und Sachsen-Anhalt und Thüringen in Teilen im Vergleich zu den anderen Bundesländern festgestellt werden. Tendenziell waren jedoch alle regionalen Bildungsrenditen in den ostdeutschen Bundesländern höher als in den westdeutschen Bundesländern. Der Vergleich der westdeutschen Bundesländer untereinander zeigt Homogenität.

Auch bei der regionalen Betrachtung wurde der Datensatz um die Erwerbslosen erweitert. Hierfür wurde das Heckman-Verfahren zunächst über einen nach Bundesländern separierten Datensatz angewendet. Es ergaben sich wiederum in der Rangfolge ähnliche Ergebnisse wie zuvor bei der Schätzung über das 3SLS-Verfahrens bzw. OLS-Verfahren mit Interaktionstermen. Nachteilig bei dieser Anwendung war, dass durch die Separierung das Heckman-Verfahren für einige Bundesländer nicht durchführbar war, da die Gruppe der Arbeitslosen teilweise einen zu geringen Stichprobenumfang hatte. Dafür wurde in einem zweiten Schritt das Modell wiederum mit Interaktionstermen im Rahmen des Heckman-Verfahrens geschätzt. Durch die gesamtdeutsche Betrachtung ergab sich hierbei nicht das Problem des zu geringen Stichprobenumfangs bei einzelnen Bundesländern. Es konnte demnach für alle Bundesländer eine einzelne Bildungsrendite geschätzt werden.

7 Schlussbetrachtung

Der Vergleich dieser Ergebnisse bestätigt ein weiteres Mal die signifikant deutlich höhere Bildungsrendite von Mecklenburg-Vorpommern – auch an dieser Stelle wurde ein Signifikanztest über den Wald-Test durchgeführt. Auch die tendenziell höheren Renditen von Sachsen-Anhalt und Thüringen wurden bestätigt. Demnach zeigte sich, dass sich unabhängig von Modell und Verfahren die Reihenfolgen der ermittelten regionalen Bildungsrenditen in gleicher Weise unterscheiden. Die absoluten Höhen variierten je nach Methode. So wurde, wie in der Literatur üblich, eine deutlich höhere Rendite mit dem 3SLS-Verfahren geschätzt, als es das Heckman- oder das OLS-Verfahren zeigten. Das Ziel dieser Arbeit war es im Besonderen, die regionalen Unterschiede aufzudecken.

Aufgrund vielfältiger Einflüsse auf den Lohn und die Bildungsdauer war es nicht das vorrangige Ziel, eine absolute Höhe der Bildungsrendite festzustellen, sondern lediglich das Intervall für eine durchschnittlich jährliche Verzinsung sollte gezeigt werden. Im besonderen Fokus standen die regionalen Unterschiede, die unabhängig vom Verfahren und von der Methode nachweislich sind.

Die sich anschließende Diskussion über die möglichen Ursachen und Auswirkungen der regionalen Bildungsrenditen zeigte im Vergleich, dass es einen tendenziell negativen Zusammenhang zwischen dem mittleren Einkommensniveau und der Bildungsrendite gibt. Ein geringes Einkommensniveau konnte demnach in diesen Bundesländern für Geringqualifizierte gezeigt werden. Aufgrund dessen war von einem tendenziell geringeren Preisniveau – näherungsweise über die durchschnittlichen Baulandpreise abgebildet – in Ländern mit hohen Bildungsrenditen auszugehen. In der Analyse bestätigt sich diese Tendenz. Weiterhin ergab sich – den Erwartungen entsprechend – eine große relative Abweichung der Einkommen bei hohen Bildungsrenditen. Betrachtete man Wanderungsbewegungen je nach Qualifikation, so ergab sich eine deutlich höhere Abwanderungsquote der Geringqualifizierten aus Ländern mit einer hohen Bildungsrendite als der Hochqualifizierten. Zum einen ist das auf die höhere Notwendigkeit wegen eines höheren Arbeitslosigkeitsrisikos zurückzuführen, auf der anderen Seite konnte vermutet werden, dass die hohe Steigerung des Einkommens durch ein zusätzliches Bildungsjahr ein Anreiz für Hochqualifizierte ist, im Land zu bleiben. Diese Diskussion kann keine abschließende Beurteilung der Einflüsse regionaler Bildungsrenditen liefern, da die Anzahl der geschätzten Bildungsrenditen, mit je einer Ausprägung pro Bundesland, zu gering ist. Es können jedoch Tendenzen abgebildet werden, die der allgemeinen Intuition entsprechen.

Diese Arbeit lieferte einen neuen Beitrag zur Erklärung regionaler Disparitäten in Deutschland mit Blick auf die Verzinsung von allgemeinem Humankapital. Dabei wurden erstma-

lig regionale Bildungsrenditen für alle Bundesländer in einer separierten und gesamtdeutschen Analyse und in vielfältiger Weise geschätzt und diskutiert. Die Ergebnisse waren mit Blick auf regionale Unterschiede stets gleich, was für ihre Konsistenz spricht. Aufbauend auf den neuen Erkenntnissen können innerdeutsche Unterschiede weiter analysiert werden. Es gilt zu diskutieren, inwieweit sich diese deutlichen Unterschiede zeitlich – gerade mit Blick auf die Überwindung der innerdeutschen Teilung – entwickelt haben. Zudem könnte durch eine mehrperiodische Betrachtung die gegebene Beobachtungszahl deutlich erhöht werden, um eventuelle Einflussgrößen auch in einer multivariaten Analyse zu diskutieren. In diesem Zusammenhang stellt sich zudem die Frage einer optimalen Bildungsrendite. Auf der einen Seite zeigt sich eine deutliche Ungleichheit bei hohen Bildungsrenditen. Auf der anderen Seite jedoch liefert die Verzinsung von allgemeinem Humankapital einen Anreiz sich zu bilden. Im Extremfall, mit einer Bildungsrendite von null, hätte kein rationales Individuum die Motivation zunächst auf Einkommen zu verzichten, um sich zu bilden. Es wurden in dieser Arbeit die rein monetären Effekte der Bildung untersucht. Neben diesen finanziellen Aspekten gibt es diverse qualitative Wirkungen auf das Individuum aber auch die Gesellschaft, die in weiteren Analysen mit einbezogen werden könnten.

Literaturverzeichnis

Acs, Z. J. und Armington, C. (2004). „The Impact of Geographic Differences in Human Capital on Service Firm Formation Rates“ *Journal of Urban Economics*, Bd. 56, N. 2, S. 244–278.

Akerlof, G. A. und Kranton, R. E. (2002). „Identity and Schooling: Some Lessons for the Economics of Education.“ *Journal of Economic Literature*, Bd. XL, S. 1167–1201.

Alonso, W. (1960). „A Theory of the Urban Land Market.“ *Paper in Regional Science*, Bd. 6, S. 149–157.

Ammermüller, A. und Dohman, D. (2004). „Individuelle und soziale Erträge von Bildungsinvestitionen.“ *Forschungsinstitut für Bildungs- und Sozialökonomie (FiBS)*, Bd. 1.

Ammermüller, A. und Weber, A. M. (2005). „Educational Attainment and Returns to Education in Germany.“ *ZEW-Discussion Papers*, Bd. 05-017.

Ammermüller, A., Kuckulenz, A. und Zwick, T. (2009). „Aggregate Unemployment decreases individual returns to education.“ *Economics of Education Review*, Bd. 28, N. 2, S. 217–226.

Anger, C., Plünnecke, A. und Schmidt, J. (2010). „Bildungsrenditen in Deutschland – Einflussfaktoren, politische Optionen und volkswirtschaftliche Effekte.“ *Institut der deutschen Wirtschaft Köln*.

Anger, C., Plünnecke, A. und Tröger, M. (2007). „Renditen der Bildung – Investitionen in den frühkindlichen Bereich.“ *Institut der deutschen Wirtschaft Köln*.

Anger, S. und Heineck, G. (2010). „Cognitive abilities and earnings – first evidence for Germany.“ *Applied Economics Letters*, Bd. 17, N.7, S. 699–702.

Anger, S. und Lupo, K. (2007). „Bildungsrenditen von Vollzeitbeschäftigten in Deutschland: Der Osten hat aufgeholt.“ *DIW-Wochenbericht*, Bd. 10, S. 149–157.

Angrist, J. und Krueger, A. B. (2001). „Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiment.“ *Journal of Economic Perspectives*, Bd. 15, N. 4, S. 69–85.

Ashenfelter, O., Harmon, C. und Oosterbeek, H. (1999). „A review of estimates of the schooling/ earnings relationship, with test for publication bias.“ *Labor Economics*, Bd. 6, N. 4, S. 453–470.

Ashenfelter, O. und Krueger, A. (1994). „Estimates of the Economic Return to Schooling from a New Sample of Twins.“ *American Economic Review*, Bd. 84, N. 5, S. 1157–1173.

Backhaus, K., Erichson, B., Plinke, W., Weiber, R. (2006 (11. Auflage)). *Multivariate Analysemethoden*. Springer Verlag.

Basmann, R. L. (1960). „On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics.“ *Journal of American Statistical Association*, Bd. 55, N. 292, S. 650–659.

Battu, H., Belfield, C. und Sloane, P. J. (1999). „Overeducation Among Graduates: a cohort view.“ *Education Economic*, Bd. 7, N. 1, S. 21–38.

Becker, G. (1974). „A Theory of Marriage.“ *National Bureau of Economic Research*, S. 299–351.

Becker, G. (1994 (3. Auflage)). *Human Capital – A Theoretical and Empirical Analysis with Special Reference to Education*. The University of Chicago Press.

Beeson, P. E. und Eberts, R. W. (1989). „Identifying Productivity and Amenity Effects in Interurban Wage Differentials.“ *Review of Economics and Statistics*, Bd. 71, N. 3, S. 443–452.

Belzil, C. und Hansen, J. (2002). „Unobserved Ability and the Return to Schooling.“ *Econometrica*, Bd. 70, N. 5, S. 2075–2091.

Benford, F. (1938). „The Law of Anomalous Numbers.“ *Proceedings of the American Philosophical Society*, Bd. 78, N. 4, S. 551–572.

Bennett, R., Glennerster, H. und Nevison, D. (1995). „Regional Rates of Return to Education and Training in Britain.“ *Regional Studies*, Bd. 29, N. 3, S. 279–295.

Berthold, N. und Müller, A. (2010). „Regionale Disparitäten in Deutschland – Auf dem Weg zu gleichwertigen Lebensverhältnissen?“ *Wirtschaftsdienst*, Bd. 90, N. 9, S. 591–597.

- Blackburn, M. L. und Neumark, D. (1995). „Are OLS Estimates of the Return to Schooling Biased Downward? Another Look.“ *Review of Economics and Statistics*, Bd. 77, N. 2, S. 217–230.
- Blundell, R., Dearden, L., Goodman, A., Reed, H. (2000). „The Returns to Higher Education in Britain: Evidence From a British Cohort.“ *The Economic Journal*, Bd. 110, N. 461, S. 82–99.
- Bogai, D., Seibert, H. und Wiethölder, D. (2007). „Pendlerbericht Berlin-Brandenburg 2006 – Arbeitskräftemobilität in den gemeinsamen Branchenkompetenzfeldern Berlin Brandenburgs.“ *IAB regional*, Bd. 2.
- Boockmann, B. und Steiner, V. (2006). „Cohort effects and the return to education in West Germany.“ *Applied Economics*, Bd. 38, N. 10, S. 1135–1152.
- Borjas, G. J. (1996). *Labor Economics*. The McGraw-Hill Companies, Inc.
- Borjas, G. J., Bronars, S. G. und Trejo, S. J. (1992). „Self-selection and internal migration in the United States.“ *Journal of Urban Economics*, Bd. 32, N. 2, S. 159–185.
- Bound, J., Jaeger, D. A. und Baker, R. M. (1995). „Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak.“ *Journal of the American Statistical Association*, Bd. 90, N. 430, S. 443–450.
- Brown, C. und Medoff, J. L. (1989). „The Employer Size-Wage Effect.“ *Journal of Political Economy*, Bd. 97, N. 5, S. 1027–1059.
- Bundesamt, S. (2008a). „Bildungsfinanzbericht 2008.“ *Destatis*.
- Bundesamt, S. (2008b). „Regionalstatistik“, <https://www.regionalstatistik.de/genesis/online/-online;jsessionid=D024E1173F2754DE7A1C0AA3A06CCD28>, Abruf: 13.08.2010.’ *Destatis*.
- Card, D. (1995). „Using Geographic Variation in College Proximity to Estimate the Return to Schooling.“ *Aspects of Labour Economics: Essays in Honour of John Vanderkamp*.
- Card, D. (1999). „The causal effect of education on earnings.“ *Handbook of Labor Economics*, Bd. 3, S. 1801–1863.
- Card, D. (2001). „Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems.“ *Econometrica*, Bd. 69, N. 5, S. 1127–1160.

Card, D. und Lemieux, T. (2001). „Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis.“ *Quarterly Journal of Economics*, Bd. 116, N. 2, S. 705–746.

Chiswick, B. R. (1968). „The Average Level of Schooling and the Intra-Regional Inequality of Income: A Clarification.“ *American Economic Review*, Bd. 58, N. 3, S. 495–500.

Chiswick, B. R. (2006). „Jacob Mincer, Experience and the Distribution of Earnings.“ *Jacob Mincer: A Pioneer of Modern Labor Economics*, S. 109–126.

Chow, G. C. (1960). „Tests of Equality Between Sets of Coefficients in Two Linear Regressions.“ *Econometrica*, Bd. 28, N. 3, S. 591–605.

Christensen, B. (2001). „The Determinants of Reservation Wages in Germany. Does a Motivation gap exist?“ *Kiel working paper*, Bd. 1024.

Christensen, B. (2002). „Reservation Wage, offered Wages, and Unemployment Duration – New Empirical Evidence.“ *Kiel working paper*, Bd. 1095.

Christensen, B. (2003). „Anspruchslohn und Arbeitslosigkeit in Deutschland.“ *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, Bd. 04, S. 573–598.

Cohn, E. und Addison, J. (1998). „The Economic Returns to Lifelong Learning in OECD Countries.“ *Education Economic*, Bd. 6, N. 3, S. 253–207.

Combes, P. P., Duranton, G. und Gobillon, L. (2008). „Spatial wage disparities: Sorting matters!“ *Journal of Urban Economics*, Bd. 63, N. 2, S. 723–742.

Cragg, J. (1967). „On the Relative Small-Sample Properties of Several Structural- Equations Estimators.“ *Econometrica*, Bd. 35, N. 1, S. 89–110.

Dechênes, O. (2007). „Estimating the Effects of Family Background on the Return to Schooling.“ *Journal of Business & Economic Statistics*, Bd. 25, N. 3, S. 265–277.

Denny, K. und Harmon, C. (2000). „Education policy reform and the return to schooling from instrumental variables.“ *University College Dublin*, Bd. 4.

Deutsches Institut für Wirtschaftsforschung (2011). „SOEP Dokumentation.“ [Http://www.-diw.de/deutsch/soep/26628.html](http://www.diw.de/deutsch/soep/26628.html), Abruf: 25.07.2011.

Dhrymes, R. J. (1969). „Alternative Asymptotic Tests of Significance and Related Aspects of 2SLS and 3SLS Estimated Parameters.“ *The Review of Economic Studies*, Bd. 36, N. 2, S. 213–226.

- DiNardo, J. E. und Pischke, J. (1997). „The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?“ *Quarterly Journal of Economics*, Bd. 112, S. 291–303.
- Durbin, J. (1954). „Errors in Variables.“ *Review of the International Statistical Institute*, Bd. 22, N. 1/3, S. 23–32.
- Eckey, H. F., Kosfeld, R. und Dreger, C. (2004). *Ökonometrie: Grundlagen – Methoden – Beispiele*. Gabler Verlag (3. Auflage).
- Fersterer, J. und Winter-Ebmer, R. (2003). „Are Austrian returns to education falling over time?“ *Labour Economics*, Bd. 10, N. 1, S. 73–89.
- Fertig, M. (2005). „Berufliche Qualifizierung – Umbau oder Abbau der Förderung?“ Vortrag vom Reinisch Westfälischen Institut für Wirtschaftsforschung Essen.
- Gebel, M. und Pfeiffer, F. (2007). „Educational Expansion and its Heterogenous Returns for Wage Workers.“ *ZEW Discussionpaper*, Bd. 10.
- Goebel, J., Frick, J. R. und Grabka, M. M. (2009). „Preisunterschiede mildern Einkommensgefälle zwischen West und Ost.“ *Wochenbericht des DIW*, Bd. 51-52, S. 888–894.
- Granato, N. und Niebuhr, A. (2009). „Verluste in Ostdeutschland gehen zurück.“ *IAB-Kurzbericht*, Bd. 7.
- Greene, W. H. (2008 (6. Auflage)). *Econometric Analysis*. Pearson International Education.
- Griliches, Z. (1977). „Estimating the Returns to Schooling: Some Econometric Problems.“ *Econometrica*, Bd. 45, N. 1, S. 1–22.
- Gronau, R. (1974). „Wage Comparisons – A Selectivity Bias.“ *Journal of Political Economy*, Bd. 82, N. 6, S. 1119–1143.
- Hackl, P. (2007). *Einführung in die Ökonometrie*. Pearson Studium.
- Hammen, A. (2008). „Fachliche Zusammensetzung von Bildungsportfolios.“ *Institut für Arbeitsrecht und Arbeitsbeziehungen in der EG*, Bd. 2.
- Harmon, C. und Walker, I. (1995). „Estimates of the Economic Return to Schooling for the United Kingdom.“ *American Economic Review*, Bd. 85, N. 5, S. 1278–1286.
- Haupt, A. und Janeba, E. (2003). „Bildung im Zeitalter mobilen Humankapitals.“ *Vierteljahreshefte zur Wirtschaftsforschung*, S. 173–187.

- Hausman, J. A. (1978). „Specification Tests in Econometrics.“ *Econometrica*, Bd. 49, N. 6, S. 1251–1271.
- Heckman, J. J. (1976). „The common structure of statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a simple Estimator for such Models.“ *Annals of Economic and Sozial Measurement*, Bd. 5, N. 4, S. 475–492.
- Heckman, J. J. (1979). „Sample Selection Bias as a Specification Error.“ *Econometrica*, Bd. 47, N. 1, S. 153–161.
- Heckman, J. J., Lochner, L. und Todd, P.E. (2006). „Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond.“ *Handbook of the Economics of Education*, Bd. 1, S. 307–458.
- Hedges, L. V. (1992). „Modeling Publication Selection Effects in Meta-Analysis.“ *Statistical Science*, Bd. 7, N. 2, S. 246–255.
- Heij, C., De Boer, P., Franses, P. H. et al. (2004). *Econometric Methods with Applications in Business and Economics*. Oxford University Press.
- Heinicke, K. und Thomsen, S. (2011). „Qualifikation und Fähigkeiten: Ein empirischer Vergleich von Sachsen-Anhalt mit ausgewählten Bundesländern.“ *Universität Magdeburg – working paper*, Bd. 16.
- Hámori, S. (2007). *Essays in Labour Economics and Economics of Education*. Inauguraldissertation der Universität Mannheim.
- Hufnagel, R. (2001). „Zur Berücksichtigung des Alters in Mincers Bruttostundenlohnfunktion.“ *Konferenz für SAS-Anwender in Forschung und Entwicklung*.
- Hunkler, C. (2010). *Ethnische Unterschiede beim Zugang zu Ausbildung und Erwerb von Ausbildungsabschlüssen*. Springer.
- IAB-Aktuell (2011). „Jeder fünfte Geringqualifizierte ist arbeitslos.“ *Bericht des Instituts für Arbeitsmarkt- und Berufsforschung*.
- Ichino, A. und Winter-Ebmer, R. (1999). „Lower and upper bounds of returns to schooling: An exercise in IV estimation with different instruments.“ *European Economic Review*, Bd. 43, N. 4-6, S. 889–901.
- Jaeger, D. (2003). „Estimating the returns to education using the newest current population survey education questions.“ *Economic Letters*, Bd. 78, N. 3, S. 385–394.

- Jaeger, D. und Parys, J. (2004). „On the Sensitivity of Returns to Schooling Estimates to Estimation Methods, Model Specification, and Influential Outliers if identification is weak.“ *Diskussionspapier der DFG-Forschergruppe Heterogene Arbeit*.
- Jochmann, M. und Pohlmeier, W. (2004). „Der Kausaleffekt von Bildungsinvestitionen: Empirische Evidenz für Deutschland.“ *ZEW Discussionpaper*.
- Jovanovic, B. und Rob, R. (1989). „The Growth and Diffusion on knowledge.“ *Review of Economic Studies*, Bd. 56, N. 4, S. 569–582.
- Juster, F. T. (1975). „Education, Income, and Human Behavior.“ *National Bureau of Economic Research*, S. 71–94.
- Koser, K. und Salt, J. (1997). „The geography of highly skilled international migration.“ *International Journal of Population Geography*, Bd. 3, N. 4, S. 285–303.
- Krenz, A. (2008). „Theorie und Empirie über den Wirkungszusammenhang zwischen sozialer Herkunft, kulturellem und sozialem Kapital, Bildung und Einkommen in der Bundesrepublik Deutschland.“ *SOEPpapers on Multidisciplinary Panel Data Research*, Bd. 128.
- Krueger, A. B. und Pischke, J. S. (1995). „A Comparative Analysis of East and West German Labor Markets: Before and After Unification.“ *National Bureau of Economic Research*, Bd. 4154, S. 405–446.
- Krueger, A. und Taylor, T. (2000). „An interview with Zvi Griliches.“ *Journal of Economic Perspectives*, Bd. 14, N. 2, S. 171–189.
- Krugman, P. (1994). „Past and Prospective Causes of High Unemployment.“ *Economic Review*, S. 24–43.
- Kuznet, S. (1955). „Economic Growth and Income Inequality.“ *American Economic Review*, Bd. 45, N. 1, S. 1–28.
- Lauer, C. (2000). „Enrolments in Higher Education in West Germany: The Impact of Social Background, Labour Market Returns and Educational Funding.“ *ZEW Discussionpapers*, Bd. 59.
- Lauer, C. und Steiner, V. (2000). „Returns to Education in West Germany – An empirical Assessment.“ *ZEW Discussionpaper Papers*, Bd. 4.
- Lauer, C. (2002a). „A Model of Education Attainment Application to the German Case.“ *ZEW Discussionpapers*, Bd. 06.

- Lauer, C. (2002b). „Participation in higher education: the role of cost and return expectations.“ *International Journal of Manpower*, Bd. 23 N. 5, S. 443–457.
- Lawrence, E. C. (1991). „Poverty and the Rate of Time Preference: Evidence from Panel Data.“ *Journal of Political Economy*, Bd. 99, N. 1, S. 54–77.
- Lewis, H. G. (1974). „Comments on Selectivity Biases in Wage Comparisons.“ *Journal of Political Economy*, Bd. 82, N. 6, S. 1145–1155.
- Linzert, T. (2001). „Sources of German Unemployment: Evidence from a Structural VAR Model.“ *ZEW Discussionpaper*, Bd. 41.
- Liu, N. P. und Wu, D. (2007). „New Explanations for the firm Size-Wage Premium.“ *Economics Bulletin*, Bd. 10, N. 2, S. 1–7.
- Lorenz, W. und Wagner, J. (1990). „A Note on the Returns to Human Capital in the Eighties: Evidence from Twelve Countries.“ *Luxembourg Income Study Working Paper Series*, Bd. 54.
- Maier, B. M., Pfeiffer, F. und Pohlmeier, W. (2004). „Returns to Education and Individual Heterogeneity.“ *ZEW Discussionpapers*, Bd. 34.
- Mendolicchio, C. und Rhein, T. (2012). „Wo sich Bildung für Frauen mehr lohnt als für Männer.“ *IAB-Kurzbericht*, Bd. 5, S. 1–8.
- Micklewright, J., Pearson, M. und Smith, S. (1990). „Unemployment and Early School Leaving.“ *The Economic Journal*, Bd. 100, N. 400, S. 163–169.
- Miller, A. R. (1966). „Migration differentials in labor force participation: United States, 1960.“ *Demography*, Bd. 3, N. 1, S. 58–67.
- Miller, P., Milvey, C. und Martin, N. (2006). „The Return to Schooling: Estimates from a Sample of young Australian twins.“ *Labour Economics*, Bd. 13, S. 571–587.
- Mincer, J. (1958). „Investment in Human Capital and Personal Income Distribution.“ *Journal of Political Economy*, Bd. 66, No. 4, S. 281–302.
- Mincer, J. (1962). „On-the-Job Training: Costs, Returns, and Some Implications.“ *Journal of Political Economy*, Bd. 70, No.5, S. 50–79.
- Mincer, J. (1970). „The Distribution of Labor Income: A Survey With Special Reference to the Human Capital Approach.“ *Journal of Economic Literature*, Bd. 8, N. 1, S. 1–26.

- Mincer, J. (1974). *Schooling, Experience and Earnings*. National Bureau of Economic Research.
- Moeller, J. und Haas, A. (2003). *Innovation Clusters and Interregional Competition, Kap. The Agglomeration Differential Reconsidered – An Investigation with German Micro Data 1984- 1997*. Springer.
- Moosmueller, G. (2004). *Methoden der empirischen Wirtschaftsforschung*. Pearson Education.
- Mátyás, L. (1999). *Generalized Method of Moments Estimation*. Cambridge University Press.
- Murray, M. (2006). „The Bad, the Weak, and the Ugly: Avoiding the Pitfalls of Instrumental Variables Estimation.“ *SSRN eLibrary*.
- Oosterbeek, H. (2000). „Editorial: the economics of over- and under-schooling.“ *Economics of Education Review*, Bd. 19, S. 129–131.
- Pereira, P. T. und Martins, P.S. (2004). „Returns to education and wage equations.“ *Applied Economics*, Bd. 36, N. 6, S. 525–531.
- Preugschat, E. (2006). „Lecture Notes for Labor Economics – Estimation of the Returns to Human Capital Investment.“
- Psacharopoulos, G. (1977). „Family Background, Education and Achievement: A Path Model of Earnings Determinants in the U.K. and Some Alternatives.“ *The British Journal of Sociology*, Bd. 28, N. 3, S. 321–335.
- Psacharopoulos, G. (2007). „Bildungseffekte auf Beschäftigung, Einkommen und Produktivität – eine europäische Perspektive.“ *Thematisches Seminar der Europäischen Beschäftigungsstrategie*.
- Psacharopoulos, G. (1994). „Returns on investment in education: A global update.“ *The World Bank*, Bd. 22, N. 9, S. 1325–1343.
- Puhani, P. A. (1997). „Foul or Fair? The Heckman correction for sample selection and its critique; a short survey.“ *ZEW Discussionpaper*, Bd. 7.
- Reilich, J. (2006). „Return to Schooling in Germany.“ *Statistische Diskussionsbeiträge Universität Potsdam*, Herausgeber Prof. Dr. Strohe, Bd. 24.

- Reilich, J. (2012). „Regionale Bildungsrenditen – von der Methode zur Rendite.“ *Jahrbuch für Wirtschaftswissenschaften*, Bd. 63, S. 84–110.
- Rosen, S. (1976). „A Theory of Life Earnings.“ *Journal of Political Economy*, Bd. 84, N. 4, S. 45–67.
- Sachs, L. und Hedderich, J. (2006 (11. Auflage)). *Angewandte Statistik – Methodensammlung mit R*. Springer.
- Sachverständigenrat (2005). „*Erfolge im Ausland – Herausforderung im Inland*.“ Jahresgutachten 2004/2005.
- Schira, J. (2005). *Statistische Methoden der VWL und BWL*. Pearson Education.
- Schöler, K. (2007). „Transformationsprozesse und neue ökonomische Geographie – Erklärungsbeiträge der Neuen Ökonomischen Geographie zur Transformation der ostdeutschen Volkswirtschaft.“ *Regionalpolitik in der Transformationszeit - Ziele, Erfahrungen, Perspektiven, Festschrift zum 70. Geburtstag von Klaus Gloede*, S. 145–190.
- Schnabel, I. und Schnabel, R. (2002). „Family and gender still matter: The Heterogeneity of Returns to Education in Germany.“ *ZEW Discussionpapers*, Bd. 67.
- Schräpler, J. P. (2011). „Benford’s Law as an Instrument for Fraud Detection in Survey Using the Data of Socio-Economic Panel (SOEP).“ *Jahrbuch für Nationalökonomie und Statistik*, Bd. 231 N. 5, 6, S. 685–718.
- Schultz, T. W. (1961). „Investment in Human Capital.“ *American Economic Review*, Bd. 51, N. 1, S. 1–17.
- Schwartz, A. (1976). „Migration, Age and Education.“ *Journal of Political Economy*, Bd. 84, N. 4, S. 701–720.
- Spence, M. (1973). „Job Market Signaling.“ *Quarterly Journal of Economics*, Bd. 87, N. 3, S. 355–374.
- Steiner, V. und Schmitz, S. (2010). „Hohe Bildungsrenditen durch Vermeidung von Arbeitslosigkeit.“ *DIW- Wochenbericht*, Bd. 5.
- Steiner, V. und Lauer, C. (2000). „Private Erträge und Bildungsinvestitionen in Deutschland.“ *ZEW Discussionpaper*, Bd. 18.

- Suter, S. (2005). „Bildung oder Begabung? Eine ökonometrische Untersuchung zu den Ursachen individueller Lohnunterschiede.“ *Nationales Forschungsprogramm Bildung und Beschäftigung*, Bd. 5.
- Tarazona, M. (2010). „Regionale Bildungsdisparitäten und Beschäftigungsentwicklung.“ *Raumforschung und Raumordnung*, Bd. 68, N. 6, S. 471–481.
- Trostel, P., Walker, I. und Woolley, P. (2002). „Estimates of the economic return to schooling of 28 countries.“ *Labour Economics*, Bd. 9, N. 1, S. 1–16.
- Van den Berg, G. (1990). „Search Behaviour, Transitions to Non-participation and the Duration of Unemployment.“ *Economic Journal*, Bd. 100, N. 402, S. 842–865.
- Voges, W. und Mauer, A. (2004). „Rente als Bildungsrendite.“ *Projekt: Methoden und Grundlagen des Lebenslagenansatzes*.
- Von der Lippe, P. und Kladroba, A. (2002). „Repräsentativität von Stichproben.“ *Marketing*, Bd. 24, S. 227–238.
- Von der Lippe, P. und Breuer, C. C. (2010). „Datengewinnung im periodischen regionalen Preisvergleich – Die Problematik der Mieten und Immobilienpreise.“ *Jahrbuch für Regionalwissenschaft*, Bd. 30, N. 2, S. 191–222.
- Wagner, G. G., Frick, J. R. und Schupp, J. (2007). „The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements.“ *SOEPpapers*, Bd. 1.
- Weber, B. A. (2003). „Bildungsfinanzierung und Bildungsrenditen.“ *Schweizerische Zeitschrift für Bildungswissenschaften*, Bd. 25, N. 3, S. 405–430.
- Weber, B. (2002). „The link between unemployment and returns to education: evidence from 14 European countries.“ *Education and Training*, Bd. 44, N. 4, S. 171–178.
- Wewel, M. C. (2006). *Statistik im Bachelor Studium der BWL und VWL*. Pearson Education.
- Willis, R. J. und Rosen, S. (1979). „Education and Self-Selection.“ *Journal of Political Economy*, Bd. 87, N. 5, S. 7–36.
- Willis, R. (1973). „A new Approach to the Economic Theory of Fertility.“ *Journal of Political Economy*, Bd. 81, S. 14–64.
- Wolter, S. C. (2002). „Bildungsökonomie – Eine Standortbestimmung.“ *Schweizerische Zeitung für Bildungswissenschaften*, Bd. 41, N. 1, S. 149–169.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. *Massachusetts Institute of Technology*.

Wu, D. (1974). „Alternative Tests of Independence between Stochastic Regressors and Disturbances: Finite Sample Results.“ *Econometrica*, Bd. 412, N. 3, S. 529–546.

Zellner, A. und Theil, H. (1962). „Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations.“ *Econometrica*, Bd. 30, N. 1, S. 54–78.

Anhang

A.1 Theoretische Herleitung der Mincer-Gleichung

Ausgehend von Gleichung (2.24) wird vereinfachend nur der mittlere Teil der Gleichung näher betrachtet werden:

$$\ln E_t \approx sr_s + \sum_{j=s}^{t-1} k_s \left(1 - \frac{x_j}{T}\right) r_x + \ln E_0.$$

Die Berufserfahrung kann formal beschrieben werden als $x_j = j - s$. Diese Zahl der Berufsjahre ergibt sich aus der Differenz der bisher in der produktiven Phase verbrachten Lebenszeit abzüglich der allgemeinbildenden Jahre. Somit kann die Gleichung umgeformt werden zu folgendem Ergebnis

$$\begin{aligned} \sum_{j=s}^{t-1} k_s \left(1 - \frac{j-s}{T}\right) r_x &= k_s r_x \left(\sum_{j=s}^{t-1} 1 - \sum_{j=s}^{t-1} \frac{j-s}{T} \right) \\ &= k_s r_x \left(t-1-s - \sum_{j=s}^{t-1} \frac{j-s}{T} \right) \end{aligned} \quad (\text{A.1})$$

Aus Vereinfachungsgründen wird von Gleichung (A.1) zunächst nur der Summenterm weiter betrachtet. Für ihn ergibt sich

$$\begin{aligned} \sum_{j=s}^{t-1} \frac{j-s}{T} &= \sum_{j=1}^{t-1-(s-1)} \frac{j+(s-1)-s}{T} \\ &= \sum_{j=1}^{t-s} \frac{j-1}{T}. \end{aligned}$$

Nach einigen Umformungen²¹⁷ ergibt sich nun

$$= \frac{(t-s)(t-s-1)}{2T}.$$

Es gilt, dass $x_t = t - s$. Daraus ergibt sich nun

$$\sum_{j=s}^{t-1} \frac{j-s}{T} = \frac{x_t(x_t-1)}{2T}. \quad (\text{A.2})$$

Fügt man nun dieses Zwischenergebnis, Ausdruck (A.2) in Ausdruck (A.1) ergibt sich

$$\sum_{j=s}^{t-1} k_s \left(1 - \frac{j-s}{T}\right) r_x = k_s r_x \left(t-1-s - \frac{x_t(x_t-1)}{2T}\right). \quad (\text{A.3})$$

Ersetzt man diesen Ausdruck nun wieder in der Ausgangsgleichung (2.24), so erhält man:

$$\ln E_t = r_s s + k_s r_x \left(t-1-s - \frac{x_t(x_t-1)}{2T}\right) + \ln E_0. \quad (\text{A.4})$$

Dieser Ausdruck wird in Abschnitt 2.2 weiterverwendet.

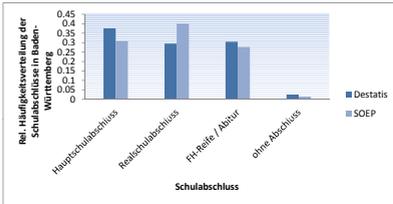
A.2 Daten und Datenbeschreibung

Dieser Anhang hat zum Ziel, die aus Gründen der Übersichtlichkeit im Hauptteil vernachlässigten Informationen der empirischen Analyse zu bündeln und der Vollständigkeit halber für den interessierten Leser zur Verfügung zu stellen. Dabei werden zunächst die Grafiken zur Prüfung der Repräsentativität pro Bundesland bereit gestellt. Im nächsten Schritt werden Teststatistiken für die Mittelwertvergleiche dargestellt um dann auf die Modellannahmen für die Regression näher einzugehen.

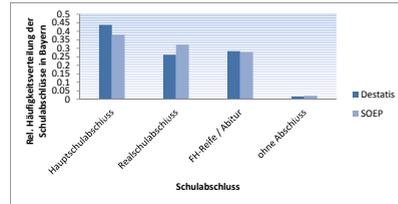
²¹⁷Anwendung findet hier die folgende allgemeine Regel: $\sum_{n=1}^N n = \frac{(N+1)N}{2}$

A.2.1 Repräsentativitätsprüfung

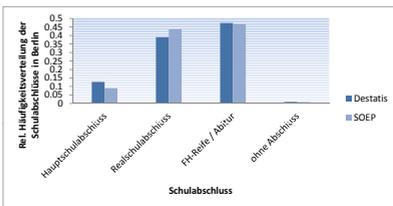
In den Abbildungen A.1 bis A.3 wird die relative Verteilung der Schulabschlüsse pro Bundesland grafisch dargestellt.



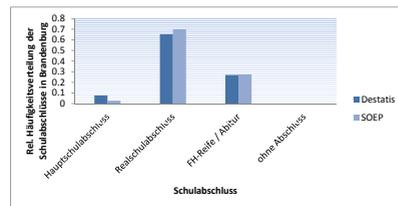
(a) Baden-Württemberg



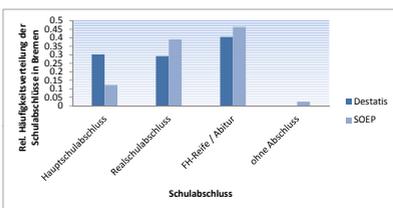
(b) Bayern



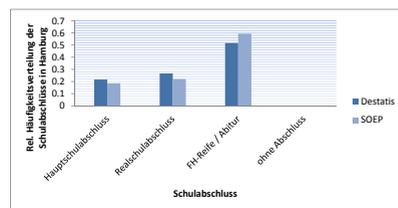
(c) Berlin



(d) Brandenburg



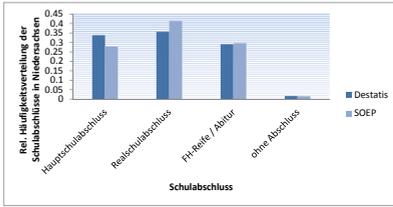
(e) Bremen



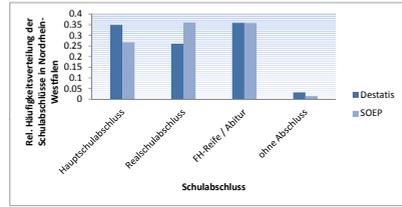
(f) Hamburg

Abbildung A.1: Verteilung der Schulabschlüsse pro Bundesland, Teil 1; Quelle: Mikrozensus, DIW, eigene Berechnungen.

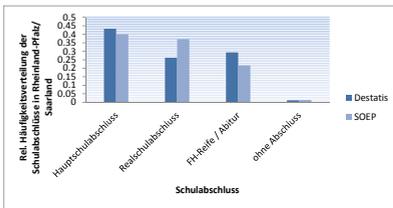
Anhang



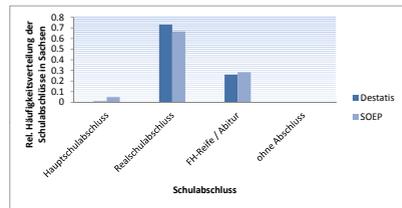
(a) Niedersachsen



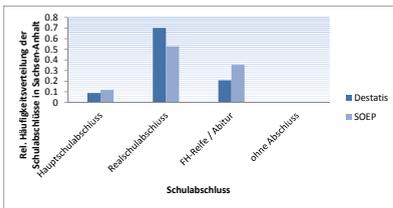
(b) Nordrhein-Westfalen



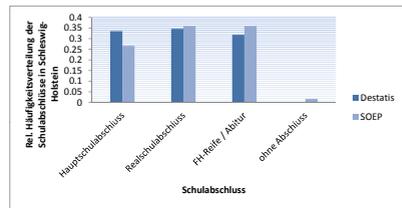
(c) Rheinland-Pfalz/ Saarland



(d) Sachsen

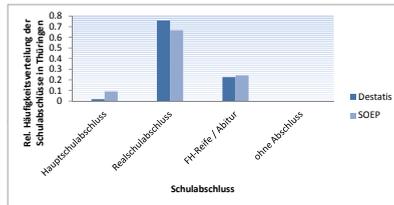


(e) Sachsen-Anhalt



(f) Schleswig-Holstein

Abbildung A.2: Verteilung der Schulabschlüsse pro Bundesland, Teil 2; Quelle: Mikrozensus, DIW, eigene Berechnungen.



(a) Thüringen

Abbildung A.3: Verteilung der Schulabschlüsse pro Bundesland, Teil 3; Quelle: Mikrozensus, DIW, eigene Berechnungen.

A.2.2 Beschreibung der Variablen und Darstellung der Testergebnisse

Zunächst werden die Testergebnisse für die Variablen zur Erklärung des Bruttostundenlohns dargestellt. Dabei liegt der besondere Fokus auf dem Einfluss durch die Bildungsdauer. Der nächste Abschnitt zeigt die möglichen erklärenden Variablen der Bildungsgleichung. Zum Schluss werden drei separierende Variablen dargestellt. Für diese Subgruppen werden im weiteren Bildungsrenditen getrennt voneinander ermittelt. Die Beschreibung der jeweils durchgeführten Tests ist in Abschnitt 4.2.1 vorgenommen. An dieser Stelle werden ausschließlich die Testergebnisse dargestellt, die nicht im Hauptteil Einzug finden und in diesem durch Verweise gekennzeichnet.

Variablen zur Schätzung der Lohngleichung

Dieser Abschnitt zeigt die Testergebnisse für die Levene-Tests und die Mittelwertvergleichstests für Variablen, die in der Lohngleichung berücksichtigt werden. Dabei wird der Einfluss der einzelnen Variablen auf den durchschnittlichen Bruttostundenlohn getestet.

Bildungsdauer. Der Fokus dieser Arbeit liegt auf dem Einfluss der Bildung auf das Einkommen. Für einen ersten Eindruck werden daher die mittleren Bruttostundenlöhne pro Bildungsdauer auf signifikante Unterschiede getestet. In Tabelle A.1 werden die P-Werte aus dem Levene-Test auf Homogenität der Varianzen dargestellt.

Anhang

Bildungsdauer	7	8,5	9	10	10,5	11	11,5	12	13	13,5	14	14,5	15	16	17	18
mittl.	28	12	251	114	804	219	1031	551	208	82	75	142	263	208	19	686
Brutto-																
stunden-																
lohn																
7	12,05	–														
8,5	15,41	0,70														
9	12,82	0,52	–													
10	13,33	0,22	0,13	–												
10,5	15,05	0,12	0,57	0,71	–											
11	16,64	0,03	0,27	0,00	0,03	–										
11,5	15,09	0,10	0,43	0,00	0,14	0,32	–									
12	15,91	0,05	0,33	0,00	0,04	0,90	0,23	–								
13	17,11	0,02	0,18	0,00	0,00	0,15	0,01	0,08	–							
13,5	18,33	0,04	0,26	0,00	0,01	0,23	0,04	0,17	0,90	–						
14	19,81	0,01	0,13	0,00	0,00	0,03	0,00	0,02	0,34	0,54	–					
14,5	17,47	0,02	0,24	0,00	0,04	0,79	0,58	0,87	0,12	0,19	0,02	–				
15	21,11	0,01	0,12	0,00	0,00	0,00	0,00	0,00	0,01	0,12	0,37	0,00	–			
16	21,52	0,00	0,10	0,00	0,00	0,02	0,00	0,00	0,41	0,66	0,68	0,02	0,06	–		
17	28,54	0,00	0,01	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,00	0,11	0,00	–	
18	26,17	0,00	0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,16	0,00	0,00	0,01	0,16	–

Tabelle A.1: P-Werte aus dem Levene-Test auf Homogenität der Varianzen pro Bildungsdauer; Quelle: DIW, eigene Berechnungen.

Aufbauend darauf ergibt sich für den Mittelwertvergleichstest die Anwendung des t- bzw. Welch-Tests. Ist der angegebene Wert kleiner als 0,05, so wird die Annahme gleicher Varianzen verworfen und somit der Welch-Test angewendet. Andernfalls kann die Nullhypothese nicht verworfen werden und es wird angenommen, dass der t-Test anwendbar ist. Die Darstellung der Ergebnisse des Mittelwerttests finden sich im Hauptteil in Tabelle 4.3.

Weitere erklärende Variablen. In Tabelle A.2 werden für die Kontrollvariablen der Lohngleichung Separierungen im Datensatz vorgenommen. Die Ergebnisse zeigen, dass alle dargestellten Variablen einen jeweils signifikant unterschiedlichen Stundenlohn aufweisen, unabhängig vom betrachteten Merkmal.

Variable	Ausprägung	mittlerer Bruttostundenlohn in €		
		Mittelwert	Levene-Test	Welch-Test bzw. t-Test
Geschlecht	männlich	19,71	0,00	0,00
	weiblich	15,37		
Familienstand	ledig (etc.)	16,38	0,00	0,00
	Ehe (etc.)	18,22		
Unternehmensgröße	klein	15,42	0,00	0,00
	groß	19,92		
ländliches Gebiet	urban	18,16	0,19	0,00
	ländlich	16,86		
Pendler	nein	16,73	0,00	0,00
	ja	18,38		

Tabelle A.2: P-Werte aus dem Vergleich des mittleren Bruttostundenlohns separiert nach einzelnen erklärenden Variablen; Quelle: DIW, eigene Berechnungen.

Variablen zur Schätzung der Bildungsdauer. In Tabelle A.3 werden für die möglichen Instrumente aus der Regression auf die Bildungsdauer Separierungen im Datensatz vorgenommen. Die Ergebnisse zeigen, dass alle dargestellten Variablen pro Kategorie jeweils signifikant unterschiedliche Bildungsdauern aufweisen.

Variable	Ausprägung	mittlere Bildungsdauer in Jahren		
		Mittelwert	Levene-Test	Welch-Test bzw. t-Test
väterliche Bildung	gering ¹	11,9	0,00	0,00
	mittel ²	13,9	0,00	0,00
	hoch ³	17,0	0,68	0,00
Anzahl der Bücher	wenig	12,1	0,00	0,00
	viel	14,3		
Abschluss DDR	nein	12,7	0,00	0,00
	ja	13,1		
Migration	nein	12,9	0,00	0,00
	ja	10,9		
Tätigkeit erlernt	nein	11,8	0,00	0,00
	ja	13,3		

Die Testergebnisse der väterlichen Bildung beziehen sich auf den Mittelwertvergleich zwischen den Gruppen mit ¹ geringer und mittlerer väterlicher Bildung, ² mittlerer und hoher väterlicher Bildung, ³ geringer und hoher väterlicher Bildung.

Tabelle A.3: P-Werte aus den Vergleichen der mittleren Bildungsdauern separiert nach einzelnen Instrumenten; Quelle: DIW, eigene Berechnungen.

Separierende Variablen. Separierte Bildungsrenditen werden für drei verschiedene Kohorten, für Personen des öffentlichen Dienstes bzw. jene in der Privatwirtschaft sowie für Voll- und Teilzeitbeschäftigte berechnet. In Tabelle A.4 sind die verschiedenen mittleren Bruttostundenlöhne und mittleren Bildungsdauern dargestellt. Ob die Ergebnisse tatsächlich unterschiedlich sind, wird wiederum über Mittelwertvergleichstests gezeigt. Die Ergebnisse der einfachen Mittelwertvergleiche für Personen im öffentlichen Dienst und in der Privatwirtschaft zeigen zunächst nicht der Intuition entsprechende Ergebnisse. Dies deutet auf verschiedene Strukturen in den beiden Sektoren. Um dies zu testen werden drei aussagekräftige Bildungsdauern ausgewählt und verglichen.

Variable	Ausprägung	mittlerer Bruttostundenlohn in €		
		Mittelwert	Levene-Test	Welch-Test bzw. t-Test
Kohorten	30–40 ¹	15,97	0,00	0,00
	41–50 ²	17,88	0,00	0,00
	51–60 ³	19,05	0,00	0,00
öffentlicher Dienst	ja	18,73	0,00	0,00
	nein	17,22		
Erwerbsstatus	Teilzeit	15,65	0,09	0,00
	Vollzeit	18,28		
		mittlere Bildungsdauer in Jahren		
Kohorten	30–40 ¹	12,9	0,00	0,00
	41–50 ²	12,6	0,02	0,30
	51–60 ³	12,8	0,00	0,09
öffentlicher Dienst	ja	13,6	0,00	0,00
	nein	12,4		
Erwerbsstatus	Teilzeit	12,7	0,98	0,74
	Vollzeit	12,8		

Die Testergebnisse beziehen sich auf den Mittelwertvergleich zwischen den Kohorten¹ 30–40 und 41–50 Jahre, ² 41–50 und 51–60 Jahre sowie ³ 30–40 und 51–60 Jahre.

Tabelle A.4: P-Werte aus den Vergleichen des mittleren Bruttostundenlohn und der mittleren Bildungsdauern separiert nach Gruppen; Quelle: DIW, eigene Berechnungen.

Tabelle A.5 zeigt die P-Werte beim Vergleich der mittleren Bruttostundenlöhne zwischen öffentlichen und privaten Arbeitnehmern pro ausgewähltem Bildungsabschluss. So wird deutlich, dass auf dem gesetzten 5 %-Niveau lediglich für Personen mit einem Hochschulabschluss ein signifikant unterschiedlicher Bruttostundenlohn berechnet werden kann.

ausgewählter Bildungsabschluss	Levene-Test	Mittelwert- vergleichs- Test
Hauptschulabschluss mit Lehre	0,01	0,49
Realschulabschluss mit Lehre	0,00	0,07
Hochschulabschluss	0,00	0,04

Tabelle A.5: P-Werte zum Vergleich mittlerer Bruttostundenlöhne ausgewählter Bildungsdauern im privaten und öffentlichen Sektor; Quelle: DIW, eigene Berechnungen.

A.3 Schätzverfahren und Ermittlung der nationalen Bildungsrendite

In diesem Abschnitt werden die aus Gründen der Übersichtlichkeit im Hauptteil ausgelassenen Schätzverfahren, Ergebnisse oder Annahmenprüfungen gezeigt. Hierbei werden im besondern die Voraussetzungen zur Durchführung der Schätzverfahren dargestellt.

A.3.1 Allgemeine Darstellung der OLS-Annahmen

Nachfolgend werden die in Abschnitt 5.1 beschriebenen Annahmen kurz zusammengefasst und die technische Umsetzung zur Prüfung sowie die Konsequenzen und mögliche Lösungen beschrieben.

Voraussetzungen zur Durchführung des OLS-Verfahrens. Die Voraussetzung für das OLS-Verfahren werden kurz zusammengefasst in den Tabellen A.6, A.7, A.8 und A.9. Dabei wird die Voraussetzung dargestellt, die Prüfung über welches Verfahren mit genauer Anwendung beschrieben sowie die Auswirkungen bei Nichterfüllung und mögliche Gegenmaßnahmen aufgezeigt. Diese Darstellungen basieren auf den Ausführungen von Backhaus et al. (2006 (11. Auflage)), Bound et al. (1995); Eckey et al. (2004); Heij et al. (2004); Wewel (2006); Wooldridge (2002).

Kriterien zur Beurteilung der OLS-Ergebnisse. Die Tabellen A.9 und A.10 zeigen aufbauend auf den Voraussetzungen des OLS-Verfahrens eine Überprüfung der Güte der gewählten Modellstruktur.

Voraussetzung	Test	Anwendung	Auswirkung bei Nichterfüllung	Lösung
V1	optische Prüfung	Grafik wobei die Linearität geprüft wird im Vergleich von abhängiger vs. unabhängiger Variablen.	<ul style="list-style-type: none"> OLS ist nicht anwendbar. 	<ul style="list-style-type: none"> Lineare Transformation.
	sinnvolle Modell-spezifikation	Ist das Modell inhaltlich sinnvoll aufgestellt?	<ul style="list-style-type: none"> Verzerrung der Schätzwerte. 	<ul style="list-style-type: none"> Modell- veränderungen durch Hinzunahme und Weglassen von Variablen.
V2	normal- verteilte Störgröße	Test auf Normalverteilung <ul style="list-style-type: none"> H_0 : Normalverteilung Teststatistik: $JB = n[\frac{1}{6}S^2 + \frac{1}{24}(K-3)^2]$ Richtwerte sind $S \sim 0, K \sim 3$. Folgt näherungsweise $\chi^2_{0,95}(2)$ – Verteilung. 	<ul style="list-style-type: none"> Statistische Tests sind nicht anwendbar. 	<ul style="list-style-type: none"> Näherungsweise kann bei großen Fallzahlen ($n > 40$) von einer Normalverteilung der Schätzer ausgegangen werden.

Tabelle A.6: Darstellung der OLS-Annahmen, Teil 1.

Voraussetzung	Test	Anwendung	Auswirkungen	Lösung
V3 Homoskedastizität	Breusch-Pagan-Test	<p>Test auf Homoskedastizität</p> <ul style="list-style-type: none"> • H_0 : Homoskedastizität. • Quadrierte Residuen der OLS-Regression werden in einer Zusatzregression als Regressand auf die Regressoren des Ursprungsmodells geschätzt. • Berechnung des Lagrange-Multiplikators über $LM_{BP} = nR^2$. • Folgt näherungsweise $\chi^2(p-1)$ – Verteilung, mit $(p-1)$ unabhängigen Variablen. 	<ul style="list-style-type: none"> • Ineffiziente Koeffizienten. 	<ul style="list-style-type: none"> • Ineffiziente Schätzer.
V4 Exogenität von Regressoren	Durbin-Wu-Hausmann-Test	<p>Test auf Exogenität der Regressoren.</p> <ul style="list-style-type: none"> • H_0 : Exogenität der Regressoren. • Vergleich der Ergebnisse aus Instrumentvariablen- und OLS-Regression • Zusatzregression: $ResOLS = \alpha + \beta x_i + ResAUX$, wobei $ResAUX$ für die geschätzten Residuen aus der IV-Regression steht. 	<ul style="list-style-type: none"> • Nicht erwartungstreue, inkonsistente und ineffiziente Schätzer 	<ul style="list-style-type: none"> • Instrumentvariablen-Schätzung

Tabelle A.7: Darstellung der OLS-Annahmen, Teil 2.

Voraussetzung	Test	Anwendung	Auswirkungen	Lösung
V4		<ul style="list-style-type: none"> • Berechnung des Lagrange-Multiplikators über $LM_{DWH} = nR^2$. • Folgt näherungsweise einer $\chi^2(k_0)$ – Verteilung, wobei k_0 die Anzahl der endogenen Variablen ist. 		
V5	Räumliche Autokorrelation	<ul style="list-style-type: none"> • Zeitliche Autokorrelation kann bei einer Querschnittsanalyse ausgeschlossen werden. • Räumliche Autokorrelation wird im Vergleich zweier Grafiken geprüft. Dabei werden die Zusammenhänge zwischen Residuen und geschätzten Bruttostundenlöhnen für Ost-Westdeutschland separat dargestellt. 	<ul style="list-style-type: none"> • Ineffiziente Schätzer. 	<ul style="list-style-type: none"> • Schätzungen für die Regionen separat.
V6	Keine Multikollinearität	<p>Berechnung der VIF für die einzelnen Regressoren</p> <ul style="list-style-type: none"> • Modell: $x_i = \alpha + \beta_j x_j + \varepsilon_i$. 	<ul style="list-style-type: none"> • Ineffiziente Schätzer. 	<ul style="list-style-type: none"> • Erweiterung des Stichprobenumfangs.

Tabelle A.8: Darstellung der OLS-Annahmen, Teil 3.

Kriterien/ Voraussetzung	Test	Anwendung	Auswirkungen	Lösung
V 6		<ul style="list-style-type: none"> Daraus ergibt sich pro Regressor ein R^2. Berechnung $VIF = \frac{1}{1-R_i^2}$. Kritischer Wert: $VIF > 20$. 	<ul style="list-style-type: none"> Das Bestimmtheitsmaß wird künstlich erhöht. Der t-Test ist oft nicht valide. Die Koeffizienten reagieren stark auf Modellveränderungen. 	<ul style="list-style-type: none"> Nicht notwendige Regressoren eliminieren. Faktorenbildung aus den multikollinearen Variablen. Akzeptieren.
K1	$korr: R^2$	<p>Anteil der erklärten Streuung an der gesamten Streuung.</p> <ul style="list-style-type: none"> $R^2 = \frac{\text{erklärte Streuung}}{\text{gesamte Streuung}}$ Wird durch die Hinzunahme von Regressoren automatisch erhöht, auch wenn diese redundant sind. $R_{korr}^2 = R^2 - \frac{J(1-R^2)}{n-J-1}$, wobei J die Anzahl der unabhängigen Variablen ist Intervall: $0 \leq R^2 \leq 1$ 	<ul style="list-style-type: none"> Je höher das R^2 umso besser kann die Streuung der abhängigen Variable erklärt werden. 	<ul style="list-style-type: none"> Bei zu geringem R^2 sollten weitere Regressoren hinzugezogen werden.

Tabelle A.9: Darstellung der OLS-Annahmen, Teil 4.

Kriterien	Test	Anwendung	Auswirkungen	Lösung
K1 Signifikanz des Gesamtmodells.	F-Test	<p>Liefert das Regressionsmodell einen Erklärungsbeitrag für den Bruttostundenlohn?</p> <ul style="list-style-type: none"> • $H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = 0$ • Vergleich mit der Testvariable: $F = \frac{n-J}{J-1} \frac{R^2}{1-R^2}$ • Folgt einer F-Verteilung mit $F\left(\frac{n-J}{J}, n-J-1\right)$. 	<ul style="list-style-type: none"> • Wird H_0 nicht abgelehnt, liefert das aufgestellte Modell keinen Erklärungsbeitrag. 	<ul style="list-style-type: none"> • Anderes Modell aufstellen.
K2 Signifikanz des j-ten Regressors.	t-Test	<p>Liefert die einzelne Variable einen Erklärungsbeitrag für den Bruttostundenlohn?</p> <ul style="list-style-type: none"> • $H_0 : \beta_j = 0$ • Vergleich mit der Testvariable: $t_j = \frac{\beta_j}{s_j}$ • Folgt einer t-Verteilung. 	<ul style="list-style-type: none"> • Bei Nichtablehnung der Nullhypothese kann kein signifikanter Einfluss festgestellt werden. 	<ul style="list-style-type: none"> • In begründeten Fällen, sind solche Variablen trotzdem nicht immer aus dem Modell zu entfernen.

Tabelle A.10: Darstellung der OLS-Annahmen, Teil 5.

A.3.2 Varianzinflationsfaktoren aus der OLS-Regression

In Tabelle A.11 sind die Varianzinflationsfaktoren zur Prüfung der Multikollinearität für die jeweiligen OLS-Modellspezifikationen dargestellt.

Variable	Modell 1	Modell 2	Modell 3	Modell 4
Bildungsjahre	1	1	1	1
Erfahrung	29	29	29	29
Erfahrung ²	29	29	29	29
weiblich		3	3	3
verheiratet		2	2	2
IAT verheiratete Frau		4	4	4
Unternehmen mit mehr als 200 Mitarbeitern			1	1
ländliches Gebiet				1
Pendler				1

Tabelle A.11: Varianzinflationsfaktoren aus der OLS-Regression; Quelle: DIW, eigene Berechnungen.

A.3.3 Allgemeine Darstellung der IV-Annahmen

Zur Prüfung der Durchführbarkeit des IV-Ansatzes sowie der Überprüfung der Instrumente zur Schätzung der Bildungsgleichung sind die notwendigen Annahmen in den Tabellen A.12 und A.13 zusammengefasst.

A.3.4 Multikollinearitätsprüfung für die Reduced-Form-Regression

Die Instrumente dürfen ebenfalls nicht korreliert sein miteinander. Demnach werden in Tabelle A.14 die Varianzinflationsfaktoren auch für die Modellspezifikationen der Reduced-Form-Regression gezeigt.

Voraussetzung	Test	Anwendung	Auswirkungen	Lösung
IV – V1 Endogener Einfluss kann eliminiert werden	F-Test auf Residuum	<p>Liefert das Regressionsmodell einen Erklärungsbeitrag für das Residuum der Lohngleichung?</p> <ul style="list-style-type: none"> Durchführung einer Zusatzregression: $\varepsilon_i = \sum_{j=1}^k \xi_j Z_{ij} + \sum_{j=k}^m \xi_j X_{ij},$ wobei Z_{ij} die k verwendeten Instrumente und X_{ij} die $m - k$ Kontrollvariablen beschreibt. F-Test mit: $H_0 : \xi_1 = \xi_2 = \xi_3 = \dots = 0$ Wird die Nullhypothese abgelehnt, so liefert das Modell einen Erklärungsbeitrag für das Residuum. 	<ul style="list-style-type: none"> Der endogene Faktor konnte nicht eliminiert werden, daher sind die Schätzer weiter verzerrt. 	<ul style="list-style-type: none"> Andere Modellspezifikation.
IV – V2 Relevanz der Instrumentvariablen	F-Test	<p>Liefert das Regressionsmodell einen Erklärungsbeitrag für die Bildungsdauer?</p> <ul style="list-style-type: none"> $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \dots = 0$ 	<p>Bildungsdauer kann nicht erklärt werden.</p>	<p>Andere Modellspezifikation.</p>

Tabelle A.12: Darstellung der IV-Annahmen, Teil 1.

Voraussetzung	Test	Anwendung	Auswirkungen	Lösung
IV – V2		<ul style="list-style-type: none"> Ist die Nullhypothese abzulehnen, liefern die Instrumente einen Erklärungsbeitrag für die Bildungsdauer. 		
	R^2	Bestimmtheitsmaß der Reduced-Form-Regression zeigt die erklärte Streuung an der Gesamtstreuung.		
IV – V3	VIF	Nähere Beschreibung und die Ergebnisse finden sich in Abschnitt A.3.1.		
IV – V4	Vergleich der Variablen	Anzahl der endogenen Variablen muss kleiner oder gleich der Anzahl der Instrumente sein.	<ul style="list-style-type: none"> Das Modell ist nicht schätzbar. 	<ul style="list-style-type: none"> Wahl einer anderen Modellspezifikation.
IV – V5	Korrelationskoeffizient	Nähere Beschreibung zu Korrelationskoeffizienten sind in Abschnitt 4.3 dargestellt.	<ul style="list-style-type: none"> Die Residuen beider Regressionen sind nicht orthogonal zueinander. 	<ul style="list-style-type: none"> 3SLS-Verfahren.

Tabelle A.13: Darstellung der IV-Annahmen, Teil 2.

Instrument	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5	Modell 6
mittlere Bildung des Vaters	1	1	1	1	1	1
hohe Bildung des Vaters	1	1	1	1	1	1
Anzahl der Bücher		1	1	1	1	1
Anzahl der Schwestern			1	1	1	1
Anzahl der Brüder			1	1	1	1
Staatsangehörigkeit nicht deutsch				1	1	1
Tätigkeit im erlerten Beruf					1	
Schulabschluss aus der DDR						1

Tabelle A.14: Varianzinflationsfaktoren der Reduced-Form-Regression aus der IV-Regression; Quelle: DIW, eigene Berechnungen.

Es wird deutlich, dass für alle unabhängigen Instrumente von keinem Multikollinearitätsproblem auszugehen ist, da der Wert stets bei 1 liegt.

A.3.5 Ergebnisse der Reduced-Form-Regression für verschiedene Kohorten

Bei der Separierung des Datensatzes nach verschiedenen Altersgruppen ergeben sich für die Schätzung der Reduced-Form-Regression die in Tabelle A.15 dargestellten Koeffizienten. Dabei sind alle Koeffizienten mit einer Irrtumswahrscheinlichkeit von 5 % signifikant von null verschieden. Weiterhin zeigen die Einflüsse die erwarteten Vorzeichen und sind im Vergleich der Altersgruppen von der Höhe recht ähnlich. Die jüngste Kohorte hat eine leicht höhere Konstante, was darauf hindeutet, dass im Vergleich die jüngsten mehr Bildungsjahre erzielen. Bemerkenswert ist weiterhin, dass für die älteste Kohorte das mit Abstand höchste R^2 mit 45 % erklärter Streuung gemessen werden kann. Daher scheint es, dass für die mittlere und jüngere Kohorte noch unberücksichtigte Einflussmöglichkeiten auf die Bildungsdauer gibt.

Bildungsdauer	30–40 Jahre	41–50 Jahre	51–60 Jahre
mittlere Bildung des Vaters	1,154*** (0,129)	1,559*** (0,116)	2,660*** (0,139)
hohe Bildung des Vaters	4,151*** (0,249)	4,579*** (0,225)	4,260*** (0,257)
Anzahl der Bücher	1,407*** (0,146)	1,371*** (0,110)	1,785*** (0,122)
Anzahl der Schwestern	-0,159** (0,064)	-0,229*** (0,048)	-0,151*** (0,054)
Anzahl der Brüder	-0,335*** (0,065)	-0,149*** (0,046)	-0,170*** (0,053)
Staatsangeh. nicht deutsch	-1,164*** (0,143)	-0,910*** (0,234)	-1,254*** (0,261)
Konstante	13,439*** (0,280)	12,872*** (0,256)	12,861*** (0,294)
korr. R ²	0,33	0,33	0,45
Wald-Test	0,000	0,000	0,000
n	1.410	1.871	1.412

Tabelle A.15: Ergebnistabelle der Reduced-Form-Regression separiert nach Kohorten; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

A.3.6 Berücksichtigung der Erwerbslosen

Hierbei werden im ersten Schritt die Regressionen für ALG-I-Empfänger mit dem 3SLS-Verfahren durchgeführt. Aus Gründen der Übersichtlichkeit wurden die Reduced-Form-Regressionen nicht im Hauptteil dieser Arbeit berücksichtigt. Um die Vollständigkeit jedoch zu gewährleisten, werden die Schätzergebnisse für die Berücksichtigung der ALG-I-Empfänger in Tabelle A.16 gezeigt.

Bildungsgleichung	erweiterter Datensatz	Erwerbstätige	ALG-I-Empfänger
	ALG I-Modell 4a	Modell 4	Modell 4
mittlere väterl. Bildung	1,601*** (0,072)	1,735*** (0,073)	1,016* (0,524)
hohe väterl. Bildung	4,272*** (0,138)	4,380*** (0,140)	3,689*** (0,687)
Anzahl der Bücher	1,569*** (0,071)	1,527*** (0,072)	1,683*** (0,391)
Anzahl der Schwestern	-0,180*** (0,031)	-0,175*** (0,032)	0,042 (0,154)
Anzahl der Brüder	-0,215*** (0,031)	-0,215*** (0,031)	0,013 (0,176)
Staatsang. nicht deutsch	-1,115*** (0,144)	-1,126*** (0,145)	-0,740 (0,753)
Konstante	13,035*** (0,161)	13,060*** (0,162)	10,791*** (1,098)
R^2	0,35	0,36	0,35
Wald-Test	0,000	0,000	0,000
n	4.812	4.693	119

Tabelle A.16: Ergebnistabelle der Reduced-Form-Regression für den erweiterten (Erwerbstätig und ALG I) Datensatz zusammen und getrennt sowie den ursprünglichen Datensatz im Vergleich; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

A.4 Regionale Analyse

In der regionalen Analyse werden hauptsächlich Unterschiede in den Bundesländern aufgezeigt. Hierfür wird zum einen der Datensatz separiert und pro Bundesland das 3SLS-Verfahren angewendet. Zum anderen wird der gesamtdeutsche Datensatz verwendet. Die regionalen Variationen werden dann über Interaktionsterme hinzugezogen. In diesem Abschnitt werden die aus Gründen der Übersichtlichkeit weggelassenen Darstellungen der Regressionsergebnisse pro Bundesland zur Verfügung gestellt. Am Schluss werden die Korrelationskoeffizienten zwischen möglichen Einflußgrößen und der regionalen Bildungsrendite dargestellt.

A.4.1 Ergebnisse der Reduced-Form-Regression pro Bundesland

Die Tabellen A.17 und A.18 zeigen die Ergebnisse für die Reduced-Form-Regression des 3SLS-Verfahrens für die Einzelregressionen pro Bundesland. Hierbei ist zu beachten, dass sich durch die Separierung des Datensatz zum Teil nur geringe Stichprobenumfänge ergeben. Daher können einige Einflüsse der Instrumente nicht signifikant nachgewiesen werden. Tendenziell ist die Höhe und die Richtung der signifikanten Koeffizienten der Instrumente den Erwartungen entsprechend. Einzelne besonders hohe oder niedrige Ausprägungen sind über den erwähnten geringen Stichprobenumfang zu begründen.

A.4.2 Varianzinflationsfaktoren aus der regionalen OLS-Regression mit Interaktionstermen

Die Regionen werden neben der Separierung des Datensatzes auch über Interaktionsterme in einer gesamtdeutschen Regression berücksichtigt. Problematisch dabei ist, dass die Reduced-Form-Regression unteridentifiziert ist, bei 15 endogenen Variablen. Näherungsweise soll jedoch davon ausgegangen werden, dass die tendenzielle Richtung der OLS-Koeffizienten trotz Endogenitätsproblem gegeben ist.

Variable	VIF	Variable	VIF	Variable	VIF
D_{NW}	170	IAT_{MV}	30	IAT_{HB}	21
D_{BY}	133	Erfahrung	30	IAT_{HE}	20
D_{BW}	119	D_{HB}	29	IAT_{BY}	18
D_{NI}	97	Erfahrung ²	29	IAT_{NW}	18
D_{SN}	90	IAT_{BB}	28	IAT_{BW}	18
D_{HE}	84	IAT_{TH}	27	IAT_{w-verh}	4
$D_{RP,SL}$	78	IAT_{SN}	26	weiblich	3
D_{BB}	63	IAT_{ST}	25	verheiratet	2
D_{TH}	63	IAT_{SH}	25	ländlich	1
D_{ST}	62	IAT_{BE}	24	Pendler	1
D_{MV}	51	IAT_{RP-SL}	23	großes Unternehmen	1
D_{BE}	50	IAT_{NI}	23		
D_{HH}	34	IAT_{HH}	22		

Tabelle A.19: Varianzinflationsfaktoren aus der gesamtdeutschen OLS-Regression; Quelle: DIW, eigene Berechnungen.

Anhang

Instrumente	BW	BY	BE	BB	HB	HH	HE	MV
mittlere Bildung des Vaters	2,150*** (0,213)	1,682*** (0,200)	1,175*** (0,442)	1,047*** (0,319)	2,135*** (0,794)	0,994 (0,705)	1,654*** (0,260)	2,158*** (0,417)
hohe Bildung des Vaters	4,888*** (0,391)	4,628*** (0,395)	3,033*** (0,609)	2,966*** (0,588)	5,459*** (1,097)	3,701*** (0,985)	4,067*** (0,454)	2,954*** (0,895)
Anzahl der Bücher	1,211*** (0,211)	1,649*** (0,194)	2,185*** (0,406)	1,685*** (0,317)	1,026 (0,865)	2,289*** (0,650)	1,999*** (0,253)	1,483*** (0,379)
Anzahl der Schwestern	-0,189 (0,081)	-0,134 (0,086)	-0,329 (0,248)	-0,132 (0,169)	-0,316*** (0,334)	0,243 (0,422)	-0,182 (0,130)	-0,273 (0,170)
Anzahl der Brüder	-0,261*** (0,082)	-0,127 (0,080)	-0,197 (0,210)	-0,236 (0,150)	0,564** (0,285)	-0,264 (0,417)	-0,246** (0,105)	0,094 (0,180)
Staatsangh. nicht deutsch	-1,427*** (0,257)	-1,230*** (0,355)	-0,631 (1,613)	-4,490** (1,973)	12,320*** (0,577)	12,178*** (0,678)	-0,589 (0,489)	11,863*** (0,373)
Konstante	13,306*** (0,339)	12,670*** (0,412)	12,896*** (1,682)	16,984*** (1,987)	enfällt	enfällt	12,518*** (0,547)	enfällt
korr. R ²	0,42	0,35	0,39	0,27	0,53	0,39	0,40	0,38
Wald-Test	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
n	588	677	133	186	41	59	349	109

Tabelle A.17: Ergebnistabelle der Reduced-Form-Regression auf Bundesländerebene Teil 1; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; Quelle: DIW, eigene Berechnungen.

Instrumente	NI	NW	RP / SL	SN	ST	SH	TH
mittlere Bildung des Vaters	1,824*** (0,242)	1,764*** (0,172)	2,045*** (0,292)	1,600*** (0,228)	1,120*** (0,330)	1,618*** (0,440)	1,299*** (0,331)
hohe Bildung des Vaters	4,217*** (0,558)	4,509*** (0,387)	4,948*** (0,587)	3,848*** (0,396)	4,464*** (0,528)	4,556*** (0,933)	4,809*** (0,849)
Anzahl der Bücher	1,248*** (0,236)	1,458*** (0,164)	1,328*** (0,275)	1,258*** (0,244)	1,590*** (0,323)	1,330*** (0,428)	1,712*** (0,359)
Anzahl der Schwestern	-0,263*** (0,097)	-0,143** (0,071)	-0,162 (0,108)	-0,262** (0,116)	-0,259 (0,165)	0,236 (0,223)	-0,258** (0,122)
Anzahl der Brüder	-0,350*** (0,103)	-0,289*** (0,071)	-0,140 (0,107)	0,142 (0,120)	-0,164 (0,131)	-0,109 (0,237)	-0,112 (0,142)
Staatsangh. nicht deutsch	-0,879 (0,614)	-0,678** (0,311)	-1,004** (0,496)	-2,647 (1,911)	12,498*** (0,317)	-1,357 (1,068)	12,252*** (0,248)
Konstante	12,994*** (0,647)	12,717*** (0,346)	12,538*** (0,552)	14,848*** (1,920)	enfällt	12,927*** (1,156)	enfällt
korr. R ²	0,34	0,31	0,35	0,33	0,38	0,34	0,35
Wald-Test	0,000	0,000	0,000	0,000	0,000	0,000	0,000
n	412	971	310	349	195	128	186

Tabelle A.18: Ergebnistabelle der Reduced-Form-Regression auf Bundesländerebene Teil 2; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; Quelle: DIW, eigene Berechnungen.

Im direkten Vergleich zwischen OLS- und 3SLS-Verfahren konnten zwar in der Höhe unterschiedliche Bildungsrenditen, nicht aber in der Rangfolge Unterschiede gemessen werden. Zur Durchführung des OLS-Verfahrens muss ebenfalls die Voraussetzung V6 (Abschnitt 5.1) erfüllt sein. Die Prüfung, numerisch dargestellt in Tabelle A.19, zeigt eine hohe Korrelation zwischen den Länder-Dummy-Variablen. Auch der Blick auf die einzelnen Bundesländerinteraktionsterme zeigt eine Überschreitung des kritischen²¹⁸ in Höhe von 20. Die Werte für die Varianzinflationsfaktoren können jedoch sehr große Ausprägungen annehmen. Der höchste Varianzinflationsfaktor für die Interaktionsterme liegt bei 30. Es kann die Überschreitung des kritischen Wertes nicht ausgeschlossen werden, d. h. es muss von einem verzerrenden Multikollinearitätsproblem ausgegangen werden. Aber, da die Überschreitung nicht groß ist, wird angenommen, dass näherungsweise die Rangfolge der Koeffizienten bestehen bleibt.

A.4.3 Ergebnisse der separierten OLS-Regression

Dieser Abschnitt zeigt die Regressionsergebnisse für die Zusatzregression. Diese dient als Vergleichsinstrument zur Bewertung der ermittelten regionalen Bildungsrenditen aus dem 3SLS-Verfahren mit einem separierten Datensatz sowie den Ergebnissen aus dem OLS-Verfahren für einen gesamtdeutschen Datensatz. Die Ergebnisse der hier durchgeführten OLS-Regression mit einem separierten Datensatz sind in den Tabellen A.20 und A.21 dargestellt.

Bei dieser Regression ergeben sich für alle Bundesländer hoch signifikante Koeffizienten. Dabei hat Hamburg mit 7,4 % Einkommenssteigerung pro zusätzlichem Bildungsjahr die niedrigste Rendite. Wie auch in allen anderen Fällen ergibt sich für Mecklenburg-Vorpommern mit 15,3 % die höchste Verzinsung. Von den alten Bundesländern gibt es in Bremen mit 9,7 % die höchste Rendite. Die niedrigste Bildungsrendite in den neuen Bundesländern ist in Sachsen zu verzeichnen. Hier kann eine Verzinsung von 8,8 % pro zusätzlichem Bildungsjahr erzielt werden.

Ein statistischer Test zum Vergleich der einzelnen Bildungsrenditen pro Bundesland ist aus bereits beschriebenen Gründen, bei Durchführung von separierten Regressionen auf Bun-

²¹⁸Siehe 5.1, V6.

Regressoren	BW	BY	BE	BB	HB	HH	HE	MV
Bildungsdauer	0,080*** [0,004]	0,084*** [0,006]	0,100*** [0,014]	0,088*** [0,013]	0,093*** [0,017]	0,071*** [0,018]	0,079*** [0,007]	0,142*** [0,019]
Erfahrung	0,033*** [0,008]	0,050*** [0,016]	0,060*** [0,018]	-0,001 [0,019]	0,054*** [0,023]	-0,011 [0,030]	0,020* [0,011]	0,033 [0,029]
Erfahrung ²	-0,001*** [0,000]	-0,001*** [0,000]	-0,001*** [0,000]	0,000 [0,000]	-0,001* [0,000]	0,000 [0,001]	-0,000 [0,000]	-0,000 [0,001]
weiblich	-0,173*** [0,043]	-0,150*** [0,054]	-0,232*** [0,103]	-0,081 [0,105]	-0,120 [0,128]	-0,036 [0,119]	-0,098 [0,071]	-0,111 [0,138]
verheiratet	0,122*** [0,034]	0,061 [0,049]	0,213 [0,153]	0,036 [0,104]	0,048 [0,135]	0,407** [0,156]	0,054 [0,063]	0,063 [0,120]
IAT verh.	-0,140** [0,056]	-0,118* [0,064]	0,197 [0,174]	-0,077 [0,132]	-0,050 [0,207]	-0,470** [0,202]	-0,087 [0,082]	0,011 [0,159]
Frau	0,129*** [0,027]	0,139*** [0,028]	0,350*** [0,091]	0,197 [0,067]	0,117 [0,119]	0,120 [0,094]	0,199*** [0,037]	0,134 [0,083]
großes Unternehmen	-0,049* [0,028]	0,072** [0,029]	0,068 [0,131]	-0,057 [0,062]	0,105 [0,126]	-0,002 [0,184]	-0,047 [0,039]	0,059 [0,080]
ländliches Gebiet	0,064** [0,027]	0,030 [0,032]	0,073 [0,084]	0,121* [0,067]	0,005 [0,130]	0,183 [0,155]	0,124*** [0,041]	-0,061 [0,075]
Pendler	1,530*** [0,136]	1,110*** [0,189]	0,855*** [0,301]	1,233*** [0,346]	0,787 (0,467)	1,681*** [0,431]	1,479*** [0,216]	0,158 [0,573]
R^2	0,62	0,35	0,33	0,35	0,54	0,42	0,40	0,46
F-Test	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
JB-Test	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
BP-Test	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
n	588	677	133	186	41	59	349	109

Tabelle A.20: Ergebnistabelle der OLS-Regression auf Bundesländerebene, Teil 1; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; Quelle: DIW, eigene Berechnungen.

Regressoren	NI	NW	RP / SL	SN	ST	SH	TH
Bildungsjahre	0,078*** [0,007]	0,089*** [0,005]	0,085*** [0,008]	0,084*** [0,009]	0,107*** [0,014]	0,078*** [0,010]	0,110*** [0,010]
Erfahrung	0,039*** [0,011]	0,021*** [0,008]	0,040*** [0,014]	0,023* [0,012]	0,037*** [0,015]	0,043*** [0,015]	0,003 [0,019]
Erfahrung ²	-0,001** [0,000]	-0,000 [0,000]	-0,001** [0,000]	-0,000 [0,000]	-0,001** [0,000]	-0,001** [0,000]	0,000 [0,000]
weiblich	-0,212*** [0,074]	-0,181*** [0,048]	-0,186*** [0,069]	-0,039 [0,081]	-0,044 [0,115]	-0,109 [0,087]	-0,043 [0,069]
verheiratet	0,183*** [0,055]	0,080** [0,039]	0,091 [0,060]	0,040 [0,064]	0,251** [0,090]	0,032 [0,076]	0,132* [0,068]
IAT verh. Frau	-0,057 [0,086]	-0,135** [0,054]	-0,184** [0,088]	-0,098 [0,091]	-0,188 [0,132]	-0,117 [0,108]	-0,239** [0,106]
großes Unternehmen	0,156*** [0,039]	0,187*** [0,025]	0,233*** [0,042]	0,258*** [0,044]	0,284*** [0,063]	0,144*** [0,051]	0,352*** [0,052]
ländliches Gebiet	-0,050 [0,038]	-0,021 [0,028]	-0,053 [0,041]	-0,020 [0,040]	-0,050 [0,063]	-0,027 [0,055]	-0,050 [0,051]
Pendler	0,026 [0,040]	0,055** [0,024]	0,095** [0,046]	-0,023 [0,041]	0,076 [0,062]	0,021 [0,056]	0,004 [0,052]
Konstante	1,273*** [0,203]	1,421*** [0,175]	1,222*** [0,233]	1,052 [0,226]	0,453 [0,313]	1,272*** [0,263]	0,848*** [0,215]
R^2	0,36	0,40	0,44	0,38	0,45	0,42	0,50
F-Test	0,000	0,000	0,000	0,000	0,000	0,000	0,000
JB-Test	0,000	0,000	0,000	0,000	0,000	0,000	0,000
BP-Test	0,000	0,000	0,000	0,000	0,000	0,000	0,000
n	412	971	310	349	195	128	186

Tabelle A.21: Ergebnistabelle der OLS-Regression auf Bundesländerebene, Teil 2; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; Quelle: DIW, eigene Berechnungen.

desländerebene, nicht möglich. Daher werden die 95 %-Konfidenzintervalle der Bildungsrenditen pro Bundesland miteinander verglichen. Dargestellt sind sie in Tabelle A.23.

Bundesland	Bildungsrendite	Konfidenzintervall	Bundesland	Bildungsrendite	Konfidenzintervall
BW	0,080***	0,070–0,090	NI	0,078***	0,063–0,093
BY	0,084***	0,073–0,095	NW	0,089***	0,078–0,099
BE	0,100***	0,072–0,127	RP / SL	0,085***	0,070–0,100
BB	0,088***	0,060–0,117	SN	0,085***	0,067–0,102
HB	0,093***	0,058–0,129	ST	0,107***	0,080–0,134
HH	0,071***	0,035–0,108	SH	0,078***	0,059–0,098
HE	0,079***	0,065–0,093	TH	0,110***	0,089–0,130
MV	0,142***	0,105–0,180			

Tabelle A.23: 95 %-Konfidenzintervalle aus der separierten OLS-Regression auf Bundesländerebene; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %; Quelle: DIW, eigene Berechnungen.

Zur Veranschaulichung werden diese Intervalle zusätzlich in Abbildung A.4 aufbereitet. Oberhalb des Zahlenstrahls sind die Konfidenzintervalle der neuen Bundesländer, unterhalb der alten Bundesländer dargestellt. Es zeigt sich auch bei dieser Regression, dass es in den ostdeutschen Bundesländern eine größere Spanne in den Bildungsrenditen gibt. Die Tendenz der höheren Bildungsrenditen bleibt bestehen, ist jedoch deutlich geringer als es beim 3SLS-Verfahren war. Die Intervalle überschneiden sich im Vergleich von Mecklenburg-Vorpommern mit den westdeutschen Bundesländern – mit Ausnahme von Bremen und Hamburg – nicht. Auch zwischen Sachsen und Mecklenburg-Vorpommern kann Überschneidungsfreiheit der Intervalle festgestellt werden. Für alle weiteren Bundesländer gilt, dass von keinem signifikanten Unterschied in der Höhe der Bildungsrenditen ausgegangen werden kann. Im Vergleich zu den Intervallen, die sich aus der 3SLS-Regression ergeben, zeigen sich deutlich geringere Intervallbreiten. Die signifikanten Koeffizienten der Kontrollvariablen zeigen aus allen Regressionen den erwarteten Zusammenhang zum Bruttostundenlohn.

Die Güte der Regressionen liegt zwischen 33 % in Berlin und 62 % in Baden-Württemberg. Bei der Prüfung des Gesamtzusammenhangs kann auch für diese separierten Regressionen über den F-Test für alle regionalen Regressionen von der Ablehnung der Nullhypothese ausgegangen werden. Ähnlich wie bei der separierten 3SLS-Regression muss auch hier die Annahme der normalverteilten Residuen abgelehnt werden. Da jedoch der kritische

Wert von 40 Beobachtungen für alle Regionen überschritten ist, kann näherungsweise von normalverteilten Koeffizienten ausgegangen werden. Die Annahme der Homoskedastizität muss auch bei dieser Regression abgelehnt werden. Demnach wird ein weiteres Mal auf die korrigierten Standardfehler von White zurückgegriffen. Auch die Durchführung des Tests auf Exogenität der Regressoren muss – wie zuvor auch – abgelehnt werden. Demzufolge kann eine Verzerrung der Schätzer nicht ausgeschlossen werden. Diese Ergebnisse dienen jedoch lediglich als Vergleichsinstrument, um Unterschiede und Gemeinsamkeiten in der Höhe der Bildungsrendite zwischen den Regionen zu ermitteln. Damit fungiert dieses Verfahren als Indikator für die Stabilität der Relation zwischen den Bundesländern. Der Vergleich dieser Relation zeigt ähnliche Ergebnisse.

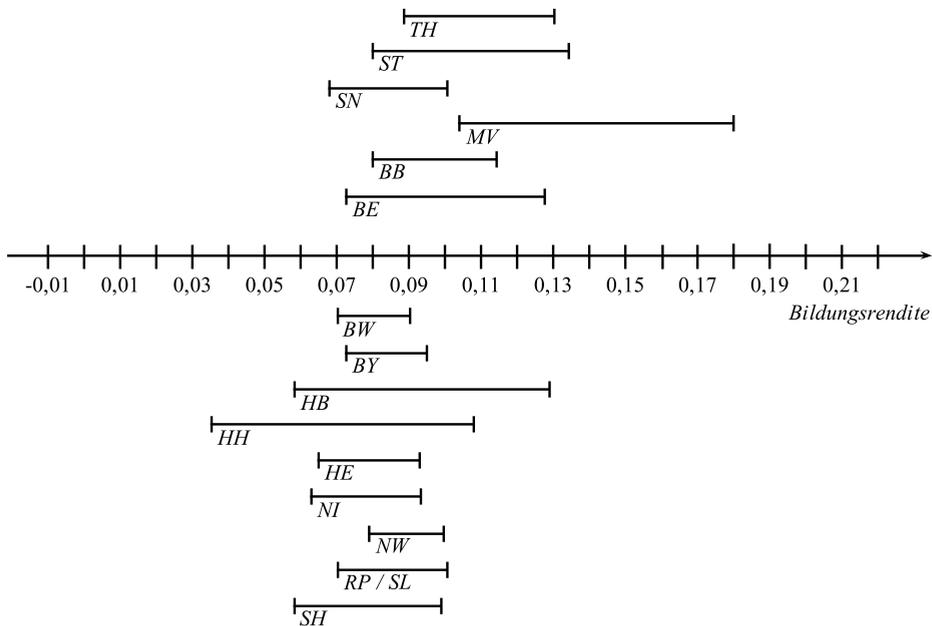


Abbildung A.4: 95 %-Konfidenzintervalle der separierten OLS-Regression auf Bundesländerebene im Zahlenstrahl; Quelle: DIW, eigene Darstellung.

Die Betrachtung eines möglichen Multikollinearitätsproblems ergibt die in Tabelle A.24 dargestellten Varianzinflationsfaktoren. Es zeigt sich, dass sie in der Höhe mit denen aus der 3SLS-Regression vergleichbar sind. Für die im Fokus dieser Arbeit stehenden Variablen, der Bildungsdauer, kann liegt kein kritisches Multikollinearitätsproblem vor.

Variable	VIF	Variable	VIF
Erfahrung	26	Bildungsjahre	1
Erfahrung ²	26	großes Unternehmen	1
IAT_{w-verh}	5	Pendler	1
weiblich	3	ländlich	1
verheiratet	2		

Tabelle A.24: VIF für die separierte OLS-Regression auf Bundesländerebene; Quelle: DIW, eigene Berechnungen.

A.4.4 Berücksichtigung der Erwerbslosen

Um alle Erwerbspersonen mit berücksichtigen zu können, muss das Verfahren von Heckman angewendet werden. Dabei wird für die regionale Analyse der Datensatz zum einen nach Bundesländern separiert, zum anderen sind die Regionen über Interaktionsterme in einer gesamtdeutschen Analyse berücksichtigt.

Berechnungen über den separierten Datensatz. Der gesamte Datensatz wird regional aufgeteilt. Pro Bundesland wird nun die Probit-Regression für das Modell aus Abschnitt 5.4.3 geschätzt und in den Tabellen A.25 und A.26 dargestellt. Zu beachten ist dabei, dass aufgrund von zu geringen Stichprobenumfängen Hamburg und Bremen nicht mit in die Analyse einbezogen werden.

Berechnung über Interaktionsterme. Die Berechnung der regionalen Bildungsrenditen auf Basis eines gesamtdeutschen Datensatzes erfolgt auch mit der Heckman-Methode über Interaktionsterme. Dabei ist die vorgelagerte Probit-Regression bereits in Abschnitt 5.4.3 dargestellt und wird für den gesamtdeutschen Raum geschätzt. Im zweiten Schritt wird die Regression in Anlehnung an das Modell in Abschnitt 6.2.2 geschätzt. Die Ergebnisse sind in Tabelle 6.23 gezeigt. In Tabelle A.27 werden die Ergebnisse für die vorgelagerte Probit-Regression gezeigt.

Regressoren	BW	BY	BE	BB	HE	MV	NI
Dauer der bisherigen Arbeitslosigkeit	-0,290*** (0,039)	-0,163*** (0,029)	-0,255*** (0,043)	-0,329*** (0,045)	-0,398*** (0,066)	-0,434*** (0,076)	-0,339*** (0,048)
Staatsangehörigkeit nicht deutsch	-0,004 (0,272)	-0,206 (0,279)	-1,838*** (0,555)	keine Variation im Datensatz	-0,115 (0,421)	keine Variation im Datensatz	-0,367 (0,489)
Anzahl der Kinder im HH	-0,004 (0,118)	-0,128 (0,089)	0,001 (0,227)	0,205 (0,179)	-0,376*** (0,138)	0,047 (0,297)	-0,048 (0,111)
Bildungsjahre	0,088* (0,048)	0,033 (0,035)	0,060 (0,054)	0,109*** (0,066)	0,101** (0,063)	0,283 (0,181)	0,031 (0,046)
weiblich	-0,578** (0,220)	-0,107 (0,171)	0,226 (0,297)	0,560*** (0,285)	-0,013 (0,269)	0,371 (0,354)	-0,505** (0,224)
verheiratet	0,326* (0,215)	0,294 (0,189)	0,086 (0,298)	0,756*** (0,277)	0,424** (0,321)	0,217 (0,348)	-0,065 (0,246)
Konstante	1,177* (0,636)	1,405*** (0,457)	0,722 (0,770)	-0,290 (0,855)	0,903 (0,822)	-1,732 (2,100)	1,977*** (0,647)

Tabelle A.25: Ergebnistabelle der Probit-Regression separiert pro Bundesland, Teil 1; Quelle: DIW, eigene Berechnungen. Signifikanzniveau:
*** 1 %, ** 5 % und * 10 %.

Regressoren	NW	RP / SL	SN	ST	SH	TH
Dauer der bisherigen Arbeitslosigkeit	-0,283*** (0,027)	-0,405*** (0,065)	-0,260*** (0,029)	-0,247*** (0,035)	-0,389*** (0,085)	-0,328*** (0,044)
Staatsangeh. nicht deutsch	0,221 (0,224)	0,047 (0,398)	-0,998 (1,096)	keine Variation im Datensatz	keine Variation im Datensatz	keine Variation im Datensatz
Anzahl der Kinder im HH	-0,001 (0,077)	-0,111 (0,132)	-0,218* (0,114)	0,071 (0,164)	-0,441* (0,234)	-0,198 (0,135)
Bildungsjahre	0,081*** (0,028)	0,008 (0,052)	0,150** (0,064)	0,061 (0,060)	0,225 (0,142)	0,002 (0,055)
weiblich	-0,108 (0,134)	-0,092*** (0,242)	0,401* (0,211)	-0,101*** (0,257)	-1,024** (0,517)	0,480* (0,252)
verheiratet	0,308** (0,140)	-0,030 (0,286)	0,655*** (0,212)	0,173 (0,263)	-0,265 (0,492)	0,041 (0,236)
Konstante	0,731** (0,372)	2,041*** (0,719)	-0,438 (0,802)	0,921 (0,823)	0,530 (1,779)	1,523** (0,724)

Tabelle A.26: Ergebnistabelle der Probit-Regression separiert pro Bundesland, Teil 2; Quelle: DIW, eigene Berechnungen. Signifikanzniveau:
*** 1 %, ** 5 % und * 10 %.

Variable	Koeffizient
Dauer der bisherigen Arbeitslosigkeit	-0,282*** (0,010)
Staatsangehörigkeit nicht deutsch	-0,108 (0,107)
Anzahl der Kinder im HH	-0,082** (0,032)
Bildungsjahre	0,064*** (0,013)
weiblich	-0,043 (0,058)
verheiratet	0,300*** (0,061)
Konstante	0,951*** (0,169)

Tabelle A.27: Ergebnisse der Probit-Regression der Heckman-Methode zur regionalen Analyse mit Interaktionstermen; Signifikanzniveau: *** 1 %, ** 5 % und * 10 %. Die Testergebnisse werden über die P-Werte gezeigt; Quelle: DIW, eigene Berechnungen.

A.4.5 Regionale Unterschiede – Zusammenhang zur Bildungsrendite

In diesem Abschnitt werden die bivariaten Korrelationskoeffizienten zur ermittelten Bildungsrendite gezeigt. Zur Berechnung dieser wird der Kendall-Tau-Korrelationskoeffizient verwandt. Dieser hat den Vorteil, dass keine Normalverteilungsannahme zugrunde liegt, was bei einer Beobachtungszahl von 15 als nicht gegeben angesehen werden kann. Zur Berechnung des Zusammenhangsmaßes werden hierbei Ränge berechnet und miteinander verglichen. Der Vorteil von Kendall-Tau gegenüber Spearman ist, dass die Differenzen zwischen den Rängen nicht gleich sein muss, da hierbei lediglich die verschiedenen Rangfolgen, nicht aber die Äquidistanz betrachtet werden. Verglichen werden nun die über das Heckman-Verfahren mit Interaktionstermen ermittelten regionalen Bildungsrenditen, mit einzelnen regionalen Größen.

Es ergibt sich hier mit einer Sicherheit von 95 % ein signifikanter Zusammenhang zwischen der Bildungsrendite und dem verfügbaren Einkommen der Haushalte pro Einwohner. Dieser Zusammenhang ist negativ. Dies bedeutet, je höher das Einkommensniveau, umso geringer ist die Bildungsrendite. Ein zweiter signifikanter, diesmal positiver Zusammenhang, besteht zwischen der Bildungsrendite und der Arbeitslosenquote. Hierbei ist

die Bildungsrendite höher, wenn auch die Arbeitslosenquote hoch ist. Der Vergleich des Variationskoeffizienten der mittleren Einkommen ergibt keinen signifikanten Zusammenhang. Allerdings ist die Korrelation mit 0,22 deutlich positiv und kann als Tendenz für einen gleichgerichteten Zusammenhang verstanden werden. Demnach geht tendenziell eine höhere Bildungsrendite mit einer höheren Abweichung der Einkommen einher.

Vergleichsvariable zur regionalen Bildungsrendite	Kendall-Tau
Verfügbares Einkommen der HH pro EW	-0,600 (0,002)
Relative Abweichung der Einkommen	0,221 (0,275)
Arbeitslosenquote	0,581 (0,003)
Durchschnittlicher Baulandpreis	-0,394 (0,087)

Tabelle A.28: Korrelationskoeffizienten zur regionalen Bildungsrendite; in Klammern sind die P-Werte für die Signifikanzniveaus angegeben.; Quelle: DIW, eigene Berechnungen.

Der Einfluss des Preisniveaus, hier näherungsweise abgebildet über die Baulandpreise, kann mit einer Sicherheit von 95 % nicht als signifikant bestätigt werden. Allerdings ergäbe sich ein anderes Ergebnis bei einer Irrtumswahrscheinlichkeit von 10 %. Für dieses Signifikanzniveau ist ein negativer Zusammenhang mit der Bildungsrendite feststellbar. Demzufolge kann auch hier die Tendenz festgestellt werden, dass ein geringeres Preisniveau tendenziell mit hoher Bildungsrendite zusammenhängt.

Kritisch ist jedoch auch bei diesen Korrelationskoeffizienten zu bemerken, dass lediglich der bivariate Zusammenhang abgebildet wird und andere Einflussgrößen nicht berücksichtigt sind.

Wie groß ist der Zusammenhang zwischen genereller Bildung und Einkommen? Die Einflüsse auf Einkommen sind vielfältig und zumeist nur schwer quantifizierbar. Daher ist die Ermittlung von einer durchschnittlichen Bildungsrendite häufig mit Schwierigkeiten verbunden. In einem methodischen Teil werden diese Probleme diskutiert und durch geeignete Verfahren (3SLS und Heckman) Lösungsmöglichkeiten aufgezeigt. Dabei werden zunächst nationale Bildungsrenditen, Renditen für verschiedene Kohorten, für Voll- und Teilzeitbeschäftigte sowie für jene die im öffentlichen oder privaten Sektor tätig sind, geschätzt.

Der zweite Teil der Arbeit zeigt die Ergebnisse regionaler Schätzungen von Bildungsrenditen. Dabei werden Unterschiede im Vergleich der Ost- und Westregion Deutschlands und auf Bundesländerebene ermittelt. Hierfür muss ein neues Regressionsmodell verwendet werden, mit dem die Möglichkeit besteht, die Gleichheit der Koeffizienten statistisch zu prüfen. Abschließend folgen eine Ursachenanalyse und die Einordnung der Arbeit, inklusive der Aufdeckung von zukünftigem Forschungsbedarf.

ISSN 2190-8702

ISBN 978-3-86956-219-3



9 783869 1562193