Sonderforschungsbereich 186 der Universität Bremen
Statuspassagen und Risikolagen im Lebensverlauf

Matthias Rasztar, Michael Windzio

„Berufsverlaufsstudie Ostdeutschland“: Feldbericht und Stichprobenevaluation der 1. und 2. Panelwelle

Sfb 186 -Arbeitspapier Nr. 67

Bremen 2000

INHALTSVERZEICHNIS

1 EINLEITUNG 3

2 GRUNDGESAMTHEIT UND STICHPROBENZIEHUNG 4

2.1 ERSTE BEFRAGUNGSWELLE 4
2.1.1 Stichprobenziehung 5
2.2 ZWEITE BEFRAGUNGSWELLE 7
2.3 METHODENWAHL 9
2.4 DURCHFÜHRUNG DER ERSTEN POSTALISCHEN BEFRAGUNG 1994 11
2.5 DURCHFÜHRUNG DER ZWEITEN POSTALISCHEN BEFRAGUNG 1997 16
2.6 BEWERTUNG DER VERSCHICKUNGSAKTION 19

3 DIE DATENBANK DER BERUFSVERLAUFSSTUDIE OSTDEUTSCHLAND 20

3.1 DAS NETTOSAMPLE NACH DER ZWEITEN ERHEBUNGSWELLE 20
3.2 PANELMORTALITÄT 21
3.3 DATENEDITION 23
3.4 DATENAUFBAU UND ORGANISATION 24
3.5 SOFTWARE ZUR DATENVERWALTUNG UND ANALYSE 25

4 STATISTISCHE EVALUATION DER STICHPROBE DER BERUFSVERLAUFSSTUDIE OSTDEUTSCHLAND 26

4.1 DAS KONZEPT DER „REPRÄSENTATIVITÄT“ 26
4.2 SYSTEMATISCHE AUSFÄLLE DURCH NON-RESPONSE UND PANELMORTALITÄT 28
4.3 WAS IST DIE GRUNDGESAMTHEIT? 29
4.4 DER χ^2 - ANPASSUNGSTEST 30
4.5 REPRÄSENTANZTEST DER BERUFSVERLAUFSSTUDIE OSTDEUTSCHLAND 32
4.5.1 Lehrabsolventen 33
4.5.2 Hochschulabsolventen 33
4.6 SCHLUßFOLGERUNG 35
4.7 ORIGINALTABELLEN OHNE ZUSAMMENFASSUNGEN 40
4.8 ORIGINALTABELLEN DER ERSTEN WELLE OHNE NACHKODIERUNGEN 41

5 LITERATUR 43

6 ANHANG 45

1 Einleitung

Mit der Öffnung der Grenze zwischen der DDR und der Bundesrepublik Deutschland im Herbst 1989, der Wirtschafts- und Währungsunion und der Vereinigung beider Staaten wurde die Transformation der DDR von einer sozialistischen Planwirtschaft zu einer sozialen Marktwirtschaft eingeleitet, deren ökonomische Folgen insbesondere auf dem Arbeitsmarkt heute noch feststellbar sind. Interessant ist dabei vor allem, wie sich die Anpassung der ökonomischen Struktur über die Allokationsmechanismen am Arbeitsmarkt und damit über die Berufsverläufe von unterschiedlichen Kohorten vollzieht, die zu unterschiedlichen historischen Zeitpunkten mit unterschiedlichen Qualifikationen in den Arbeitsmarkt eingetreten sind.

Mit der „*Berufsverlaufsstudie Ostdeutschland*“ wurde eine Erhebung von Berufsverläufen von Personen durchgeführt, die 1985, 1990 und 1995 in Rostock oder Leipzig eine berufliche Ausbildung oder ein Universitätsstudium abgeschlossen haben. Die erste Erhebung fand 1994 statt und wurde 1997 in der zweiten Panelwelle um die Absolventenkohorte 1995 ergänzt. Retrospektiv wurden mit monatsgenauen Angaben Ereignisse der beruflichen Karriere wie z.B. berufliche Wechsel oder Arbeitslosigkeitsphasen, aber auch familiäre Ereignisse wie Heirat oder Geburten erfragt.

Der vorliegende Bericht dokumentiert die Feldarbeit der Erhebung. Gegenstand von **Kapitel 2** sind die Definition der *Grundgesamtheit* und das Resultat der *Stichprobenziehung*. In den Abschnitten (2.1-2.2) werden zunächst die definierten Grundgesamtheiten und die univariaten Verteilungen von Brutto- und Nettosample der ersten und zweiten Befragungswelle dargestellt. In den Abschnitten 2.3 bis 2.7 wird die Methode der postalischen Befragung, die in der Berufsverlaufsstudie Ostdeutschland zur Anwendung gekommen ist, diskutiert und deren Durchführung in den bisherigen beiden Wellen dargestellt. **Kapitel 3** beschreibt die aus den Befragungswellen gebildete *Datenbank* der Berufsverlaufsstudie Ostdeutschland. In Abschnitt 3.1 wird das aus der zweiten Welle resultierende Nettosample dargestellt, der Abschnitt 3.2 liefert eine deskriptive Evaluation der Panelmortalität. Der Prozeß der Datenedition und die Struktur der aus mehreren Teildatensätzen bestehenden Datenbank sind Gegenstand der Abschnitte 3.3-3.4, während im letzten Abschnitt 3.5 auf die für die Auswertung von Längsschnittdaten adäquate Software hingewiesen wird. Im **Kapitel 4** wird schließlich unter Verwendung des ^{2}-Anpassungstests eine Einschätzung der Repräsentativität der Daten vorgenommen, nachdem die Problematik dieses Konzepts diskutiert wurde.

2 Grundgesamtheit und Stichprobenziehung

2.1 Erste Befragungswelle

Als Grundgesamtheit der ersten Befragungswelle der Studie wurden sämtliche Absolventen der Abschlußjahrgänge 1985 und 1990 der Universitäten Rostock und Leipzig gewählt, sowie die Absolventen einer beruflichen Lehre in den Stadtkreisen Rostock und Leipzig, die in den Jahren 1985 und 1990 abgeschlossen haben. Im Falle der Hochschulabsolventen lagen die Adressen aller Absolventen vollständig vor. Während in der uns zugänglichen Adressendatei aus Leipzig nur Hochschulabsolventen verzeichnet waren, befanden sich in der Adreßdatei der Universität Rostock sowohl die Hochschulabsolventen als auch die Studienabbrecher und Studienwechsler der Jahre 1985 und 1990 (vgl. Abb. 2.1).[1] Im Falle der Lehrlinge war der Adressenzugang durch zwei Umstände kompliziert: Einerseits wurden die Absolventenadressen nicht zentral an jeweils einer Institution in den beiden Städten aufbewahrt, andererseits waren die dezentralen institutionellen Einheiten der Adressenhaltung (Berufsschulen, Ausbildungseinheiten in Betrieben) in unterschiedlichem Ausmaß von den Umstrukturierungen im Gefolge der Wiedervereinigung betroffen. Obwohl Betriebe wurden z.T. aufgelöst und Berufsschulen z.T. zusammengefaßt wurden, gelang es uns, die Adressen der entsprechenden Kohorten über die Berufsschulen zu ermitteln. Die folgenden Adreßbestände wurden über die zuständigen Ämter Institutionen ermittelt:

Abb. 2.1.: Adreßbestand im Verhältnis zur Grundgesamtheit

Abschlußkohorte	Zahl der vorliegenden Adressen	Grundgesamtheit
Universitätsabsolventen		
Rostock 1985	1314	961
1990	1222	881
Leipzig 1985	1726	1726
1990	1597	1597
Absolventen einer beruflichen Lehre		
Rostock 1985	2431	5560
1990	2076	5589
Leipzig 1985	1399	4494
1990	2221	5621
Gesamtzahl	13986	26429

[1] Eine Bereinigung dieses Datensatzes *um die Zahl der Studienabbrecher* war vor der Datenerhebung nicht möglich. Studienabbrecher bzw. Absolventen eines anderen Jahrgangs konnten schon an der ersten Frage des Fragebogens erkennen, daß sie nicht zur Zielgruppe der Untersuchung gehörten, sie haben deshalb in der Regel nicht geantwortet. Personen, die trotzdem geantwortet haben und den Kriterien der Grundgesamtheit nicht entsprachen, wurden bei der Auswertung nicht berücksichtigt.

2.1.1 Stichprobenziehung

Aus Kostengründen, aber auch aufgrund der Tatsache, daß von Lehrlingen nicht alle Adressen vollständig vorlagen, wurde auf eine Totalerhebung der Grundgesamtheit verzichtet. Um eine ausreichende Stichprobengröße für die Wiederholungsbefragung 1997 zu erreichen, wurde eine Bruttostichprobe von ca. 8000 Personen angestrebt. Ziel der Stichprobenziehung war es, die sechs theoretisch interessanten Untersuchungsgruppen (Hochschule/Lehre; Abgangsjahr 1985/1990; Rostock/Leipzig) in etwa gleichgewichtig in die Stichprobe aufzunehmen. Für die Hochschulabsolventen der Universität Rostock bedeutete dies, daß eine Vollerhebung durchgeführt wurde. Abb. 2.2. zeigt die Verteilung dieser Gruppe auf Fachgruppen (vgl. Statistisches Bundesamt 1993). Bei den Absolventen der Universität Leipzig sollte über eine Stichprobenziehung das Bruttosample auf ca. je 1000 Absolventen der Jahre 1985 und 1990 reduziert werden. Das Sample sollte repräsentativ für die definierte Grundgesamtheit sein, eine in anderen Hochschulverbleibsstudien durchgeführte Beschränkung auf bestimmte Fachgruppen erschien uns auf der Grundlage unseres Untersuchungsziels als eine unangemessene Beschränkung der Varianz. Dennoch erscheint es plausibel, die Zugehörigkeit zu einer Fachgruppe als ein wichtiges Varianzkriterium anzusehen. In der Literatur (vgl. Böltken 1976; Kreienbrock 1989) wird in entsprechenden Situationen und bei Kenntnis der Grundgesamtheit (vgl. Merten / Teipen 1991), die Ziehung einer proportional geschichteten Zufallsstichprobe vorgeschlagen. Die Bruttostichprobe der Hochschulabsolventen der Universität Leipzig wurde mittels eines Zufallsgenerators entsprechend den Anteilen der Fachgruppen (vgl. Schnell / Hill / Esser 1989) proportional zu ihrem Vorkommen in der Grundgesamtheit (Leipziger Absolventen des entsprechenden Jahrgangs) gezogen (Abb. 2.2.).

Abb. 2.2.: Stichprobenziehung der Hochschulabsolventen

	Rostock 1985		Rostock 1990		Leipzig 1985		Leipzig 1990	
Fächergruppe	**Grundgesamtheit**	**Bruttosample**	**Grundgesamtheit**	**Bruttosample**	**Grundgesamtheit**	**Bruttosample**	**Grundgesamtheit**	**Bruttosample**
1 Sprach-/ Kulturwissenschaften	237 (18 %)	237	205 (17 %)	205	529 (31 %)	307 (31 %)	533 (33 %)	334 (33 %)
2 Sport	54 (4 %)	54	35 (3 %)	35	-	-	-	-
3 Rechts-/ Wirtschafts-/ Sozialwissenschaften	111 (8 %)	111	55 (4 %)	55	298 (17 %)	173 (17 %)	237 (15 %)	148 (15 %)
4 Mathematik, Naturwissenschaften	135 (10 %)	135	122 (10 %)	122	220 (13 %)	128 (13 %)	199 (12 %)	125 (12 %)
5 Humanmedizin	279 (21 %)	279	298 (24 %)	298	397 (23 %)	230 (23 %)	369 (23 %)	231 (23 %)
6 Veterinärmedizin	-	-	-	-	64 (4 %)	37 (4 %)	57 (4 %)	36 (4 %)
7 Agrar-/ Forst-/ Ernährungswissenschaften	251 (19 %)	251	226 (18 %)	226	161 (9 %)	93 (9 %)	164 (10 %)	103 (10 %)
8 Ingenieurwissenschaften	247 (19 %)	247	281 (23 %)	281	21 (1 %)	12 (1 %)	20 (1 %)	13 (1 %)
9 Kunst, Kunstwissenschaften	-	-	-	-	36 (2 %)	21 (2 %)	18 (1 %)	11 (1 %)
Gesamt	1314 (100%)	1314	1222 (100%)	1222	1726 (100%)	1001 (100%)	1597 (100%)	1001 (100%)

Über die Grundgesamtheit der Absolventen einer beruflichen Lehre in den Städten Rostock und Leipzig konnte in Rostock durch das Schulverwaltungsamt der Stadt eine fachliche Aufschlüsselung der Absolventen der betreffenden Jahre in Erfahrung gebracht werden[2]; in Leipzig konnte mit Hilfe des örtlichen Stadtarchives eine entsprechende Auflistung erzielt werden. Die Kenntnis dieser Daten ermöglichte aufgrund eines zentralen Merkmals von Teilgruppen der Grundgesamtheit, nämlich der Fachrichtung des Berufsabschlusses[3], über eine Schichtung der Stichprobenziehung die "Willkürlichkeit"[4] der Ziehung von Ausbildungseinheiten zu korrigieren. Abb. 2.3. zeigt, daß in Relation zur Grundgesamtheit im zugänglichen Adreßbestand eine Überrepräsentation von Fertigungsberufen im Vergleich zu Dienstleis-

[2] Wir möchten Frau Meyer für die Zusammenstellung dieser Materialien danken.

[3] Die Zuordnung von Einzelberufen zu Fachgruppen erfolgte nach der Systematik des statistischen Bundesamtes (1992).

[4] "Willkür" bezieht sich hier nicht auf die Methode unseres Vorgehens, sondern auf institutionelle Prozesse der Schließung von *Berufsschulen*, bzw. des Verbleibs ihrer Archive.

tungsberufen vorhanden war. Durch das Verfahren der proportional geschichteten Zufallsstichprobe wurde diese Disproportionalität korrigiert.[5]

Abb. 2.3.: Stichprobenziehung der Absolventen einer beruflichen Lehre [absolut und in (%)]

	Rostock 1985			Rostock 1990			Leipzig 1985			Leipzig 1990		
Fachgruppen	**Grund-gesamt-heit**	**zugäng-liche Adres-sen**	**Brutto-sample**	**Grund-gesamt-heit**	**zugäng-liche Adres-sen**	**Brutto-sample**	**Grund-gesamt-heit**	**zugäng-liche Adres-sen**	**Brutto-sample**	**Grund-gesamt-heit**	**zugäng-liche Adres-sen**	**Brutto-sample**
1 Pflanzenbauer/ Tierzüchter/ Fischereiberufe	81 (1 %)	-	-	100 (2 %)	-	-	53 (1 %)	-	-	80 (1 %)	-	-
2 Bergleute/ Mineralgewinner	-	-	-	-	-	-	-	-	-	-	-	-
3 Fertigungsberufe	3100 (56)	1573 (65)	566 (57)	2185 (49)	1109 (53)	497 (50)	3748 (67)	975 (70)	677 (68)	3437 (61)	1560 (70)	637 (64)
4 Technische Berufe	89 (2)	27 (1)	16 (2)	55 (1)	11 (1)	11 (1)	218 (4)	77 (5)	39 (4)	187 (3)	9 (0)	9 (1)
5 Dienstleistungsberufe	2290 (41)	831 (34)	418 (42)	2154 (48)	956 (46)	490 (49)	1579 (28)	347 (25)	284 (28)	1917 (34)	652 (29)	354 (35)
Gesamt	5560 (100)	2431 (100)	1000 (100)	4494 (100)	2076 (100)	998 (100)	5589 (100)	1399 (100)	1000 (100)	5621 (100)	2221 (100)	1000 (100)

2.2 Zweite Befragungswelle

Die Befragungspopulation der zweiten Welle setzte sich aus zwei Teilmengen zusammen. Die erste Teilmenge umfaßte 1667 Personen der Basiserhebung 1994, die sich 1996 mit Ihrer Unterschrift bereiterklärten, sich ein weiteres Mal befragen zu lassen. Diese Personen wurden 1997 im Laufe der Feldarbeit viermal angeschrieben, um zum einen die Adressen zu aktualisieren, und zum anderen um „Nachzufassen“, wenn die Fragebögen nicht rechtzeitig zurückgesandt wurden. Die zweite Teilmenge setzt sich aus den Absolventen des Jahres 1995 zusammen. Als Grundgesamtheit dieser Basiserhebung wurden sämtliche Absolventen der Abschlußjahrgänge 1995 der Universitäten Rostock und Leipzig gewählt sowie die Absolventen einer beruflichen Lehre in den Stadtkreisen Rostock und Leipzig, die in den Jahren 1995 abgeschlossen hatten. Im Falle der Hochschulabsolventen lagen die Adressen aller Absolventen beider Universitäten vollständig vor. Im Falle der Lehrlinge war der Adressenzugang wie schon 1994 durch die dezentrale Lagerung kompliziert: Die Absolventenadressen wurden nicht zentral an jeweils einer Institution in den beiden Städten aufbewahrt, sondern dezentral an den jeweiligen berufsbildenden Schulen beider Städte. Nach einer Datenerhebungsphase von nahezu sechs Monaten lagen jedoch alle notwendigen Anschriften vor. Aus Kostengründen wurde wie schon bei der 85er und 90er

[5] Bei den zugänglichen Adressen war die ohnehin in den Stadtkreisen nur sehr selten auftretende Gruppe von Berufen des primären Sektors nicht vertreten. In diesem Fall konnte auch durch die proportional geschichtete Zufallsstichprobe keine Korrektur in Richtung auf die Grundgesamtheit vorgenommen werden.

Kohorte auf eine Totalerhebung der Grundgesamtheit aller Absolventen des Jahres 1995 verzichtet. Um eine ausreichende Stichprobengröße für die Wiederholungsbefragung 2000 zu erreichen, wurde eine Bruttostichprobe von ca. 4000 Personen angestrebt. Ziel der Stichprobenziehung war es, die vier theoretisch interessanten Untersuchungsgruppen (Hochschule/Lehre; Rostock/Leipzig) in etwa gleichgewichtig in die Stichprobe aufzunehmen. In Rostock hatten 1.206 Personen die Universität verlassen und in Leipzig waren es 1.795. Die Abb. 2.4 und 2.5 zeigen die Verteilung der Absolventen auf die verschiedenen Fachgruppen (vgl. Statistisches Bundesamt 1993).

Abb. 2.4.: Absolventen des Jahres 1995 nach Fachbereichen der Universität Rostock

Fachbereich	Grundgesamtheit (abs.)	%	Brutto-sample (abs.)
1 Sprach-/ Kulturwissenschaften	24	2.0	20
3 Rechts-/ Wirtschafts-/ Sozialwissenschaften	122	10.1	101
4 Mathematik, Naturwissenschaften	89	7.4	74
5 Humanmedizin	201	16.7	167
7 Agrar-/ Forst-/ Ernährungswissenschaften	64	5.3	53
8 Ingenieurwissenschaften	427	35.4	354
9 Kunst, Kunstwissenschaften	11	0.9	9
10 Lehrer	268	22.2	222
	1206	100	1000

Für die Hochschulabsolventen der Universitäten Rostock und Leipzig bedeutete dies, daß das Bruttosample über die gewählte Stichprobenziehung auf je 1000 Absolventen reduziert wurde. Wie schon bei der Basiserhebung wurde die Ziehung einer proportional geschichteten Zufallsstichprobe durchgeführt. Die Bruttostichproben der Hochschulabsolventen der Universitäten Leipzig und Rostock wurden mittels eines Zufallsgenerators entsprechend den Anteilen der Fachgruppen proportional zu ihrem Vorkommen in der Grundgesamtheit (Leipziger Absolventen des entsprechenden Jahrgangs) gezogen (Abb. 2.4. und 2.5., Bruttosample).

Abb. 2.5.: Absolventen des Jahres 1995 nach Fachbereichen der Universität Leipzig

Fachbereich	Grundgesamtheit (abs.)	%	Brutto-sample (abs.)
1 Sprach-/ Kulturwissenschaften	186	10.4	104
10 Lehrer	227	12.6	126
2 Sport	88	4.9	49
3 Rechts-/ Wirtschafts-/ Sozialwissenschaften	287	16.0	160
4 Mathematik, Naturwissenschaften	447	24.9	249
5 Humanmedizin	388	21.6	216
6. Veterinärmedizin	100	5.6	56
7 Agrar-/ Forst-/ Ernährungswissenschaften	61	3.4	34
9 Kunst, Kunstwissenschaften	11	0.6	6
	1795	100	1000

Bei den Lehrabsolventen wurde über den relativen Anteil der Fachrichtung des Berufsabschlusses[6] an der Grundgesamtheit das Bruttosample gezogen. Die Sample-Bildung erfolgt mit Hilfe der Stabu-Berufsklassifikation 1992 (Feingliederung):

I	Pflanzenbauer, Tierzüchter, Fischereiberufe	0100-0699
II	Bergleute, Mineralgewinner	0700-0999
III	Fertigungsberufe	1000-5999
IV	Technische Berufe	6000-6799
V	Dienstleistungsberufe	6800-9500

Die Verteilungen der Berufsgruppen sind der Abbildung 2.6 zu entnehmen.

Abb. 2.6.: Absolventen einer beruflichen Lehre des Jahres 1995 nach Berufsgruppen

	Rostock 1995		**Leipzig 1995**	
Fachgruppen	**Grundgesamtheit (abs.)**	**Bruttosample**	**Grundgesamtheit (abs.)**	**Bruttosample**
1 Pflanzenbauer/ Tierzüchter/ Fischereiberufe	-	-	-	-
2 Bergleute/ Mineralgewinner	-	-	-	-
3 Fertigungsberufe	696 (42,3%)	423 (42,3%)	1582 (48,3%)	483 (48,3%)
4 Technische Berufe	31 (1,9%)	19 (1,9%)	22 (0,7%)	7 (0,7%)
5 Dienstleistungsberufe	919 (55,8%)	558 (55,8%)	1670 (51,0%)	510 (51,0%)
Gesamt	1646 (100)	1000 (100)	3274 (100)	1000 (100)

2.3 Methodenwahl

Als Untersuchungsmethoden zogen wir eine mündliche Befragung durch Interviewer, eine telefonische und eine postalische Befragung in Betracht. Für das Verfahren einer mündlichen Befragung sprach, daß die Komplexität eines für Retrospektiverhebungen beruflicher Verläufe notwendigen Fragebogens mit Hilfe eines Interviewers leichter zu bewältigen ist (vgl. ausführlich hierzu Brückner 1990). Gegen eine mündliche Befragung sprach, daß die Untersuchungspopulation über Gesamtdeutschland (teilweise auch im Ausland) verstreut lebt, so daß bei diesem Erhebungsverfahren enorme Kosten entstanden wären, die dem Projekt nicht zur Verfügung standen. Das Verfahren der telefonischen Befragung (Frey u.a. 1990) hätte demgegenüber, neben anderen Vorteilen wie z.B. einer besseren Interviewerkontrolle, eine Kostenersparnis gewährleistet. Nachteilig ist allerdings, daß bisher wenig Erfahrungen bei der Verwendung

[6] Die Zuordnung von Einzelberufen zu Fachgruppen erfolgte nach der Systematik des statistischen Bundesamtes (1992).

von Telefoninterviews bei Berufsverlaufserhebungen vorliegen. Bei unserer geplanten Untersuchung sprach vor allem die zu diesem Zeitpunkt nach wie vor geringere Telefondichte in der ehemaligen DDR gegen dieses Verfahren, da bekannt ist, daß die Versorgung mit Telefonanschlüssen in der DDR mit einer gewissen Staatsnähe korrelierte (und damit als nicht zufällige Varianz einzustufen ist), ebenso wie nach der Wende ein gewisser Zusammenhang von Telefonanschluß und wirtschaftlichem Erfolg zu vermuten war. Ein entsprechender durch die Methode induzierter Samplebias war bei unserer Untersuchung nicht erwünscht.

Für eine postalische Befragung in der zweiten Welle sprach überdies der Erfolg der ersten postalischen Erhebung, der geringere finanzielle Aufwand und das Wegfallen von unkontrollierbaren Interviewereinflüssen in der Erhebungssituation. Angesicht der gravierenden praktische Umsetzungsprobleme, welche die anderen Verfahren aufweisen, ist es verständlich, daß bei vergleichbaren Hochschulverbleibsstudien die Methode der postalischen Befragung zur Anwendung kam (z.B. Teichler/ Winkler 1990; Minks/ Bathke 1992). Als Nachteile einer postalischen Befragung gelten in der Methodenliteratur geringe Rücklaufquoten mit nicht-zufälligen Ausfallraten bei bestimmten Befragtengruppen, sowie erhöhte Anforderungen an den Fragebogen in bezug auf dessen Umfang sowie dessen Gestaltung, da Befragte anders als bei mündlichen oder telefonischen Befragungen nicht auf die Unterstützung durch Interviewer zurückgreifen können. In den letzten Jahrzehnten gab es umfangreiche methodische Untersuchungen zur Verbesserung des Verfahrens der postalischen Befragung (Dillman 1983; Hippler 1988, Porst 1999, Reuband 1999).

Nach einem Abwägen der Vor- und Nachteile der verschiedenen Verfahren entschieden wir uns für die Methode der schriftlichen Befragung. Entsprechend der von Dillman vorgeschlagenen "Total Design Method" wurde angestrebt, möglichst viele der in der Methodenliteratur vorgeschlagenen Methoden der Optimierung postalischer Befragungen zu berücksichtigen: Für den Fragebogen wurde ein kleines Format (DIN A5) verwendet. Der Fragebogen wurde grafisch ansprechend gestaltet. In einem Anschreiben wurde die arbeitsmarktpolitische Bedeutung der Untersuchung unterstrichen, ein Anliegen, das durch die Unterschriften der Wirtschaftsminister der Länder Sachsen und Mecklenburg-Vorpommern sowie des Präsidenten der Bundesanstalt für Arbeit betont wurde. Es wurden drei bzw. vier Verschickungen vorgenommen (Fragebogen; Erinnerungsschreiben; zweite Verschickung des Fragebogens und dritte Verschickung des Fragebogens). Als materieller Anreiz zur Teilnahme an der Befragung wurde bei der zweiten Verschickung des Fragebogens den Absolventen einer beruflichen Lehre ein "kleines Dankeschön" im Wert von 11,- DM in Briefmarken versprochen.

Da sich in der Basisbefragung 1994 die Erhebung beruflicher Verläufe in ihrer zeitlichen Struktur bewährt hatte, wurde das Erhebungsdesign in der 2. Welle beibehalten. Überdies hätte eine Änderung der Methode natürlich die Vergleichbarkeit der Panelwellen beeinträchtigt.

Der in der bereits klassischen Untersuchung der Lebensverlaufsstudie verwendete Fragebogen (Mayer/ Brückner 1989) erschien aufgrund seines Umfangs und seiner Komplexität für eine postalische Befragung nicht verwendbar. Bei den in den Hochschulverbleibsstudien verwendeten Fragebögen (Teichler/ Winkler 1990; Minks/ Bathke 1992), die sich in schriftlichen Befragungen bewährt hatten, war dagegen

der geringe Umfang und die ungenaue Erfassung von zeitlichen Strukturen des beruflichen Verlaufs mit unseren spezifischen Untersuchungszielen nicht kompatibel. Als Kompromiß zwischen diesen beiden Extremen erschien uns ein von Prof. Dr. Marlis Buchmann entwickelter Fragebogen, bei dem trotz eines relativ geringen Umfanges detailliert die Erwerbsgeschichte von jüngeren Absolventen über einen längeren Zeitraum erhoben wurde.[7] Dieser Fragebogen hatte sich bei einer postalischen Befragung in der Schweiz bewährt. Die Besonderheit der Fragebogenstruktur dieser Untersuchung bestand darin, daß in den ersten Fragen eine lückenlose Erhebung der zeitbezogenen Struktur berufsbezogener Episoden vorgenommen wurde, denen dann eine genauere Nachfrage zu den Details jeder einzelnen Berufsepisode in ihrem zeitlichen Ablauf folgte. Einstellungsfragen und Fragen zum familiären Verlauf beschlossen den Fragebogen.

Mit vielen Variationen wurde die Grundstruktur dieses Fragebogens in unserer Untersuchung übernommen (s. ein Beispiel im Anhang), da sie die Genauigkeit der Erfassung zeitlicher Verläufe mit einer leicht verständlichen (für die Befragten nachvollziehbaren) Gliederung verband. Die Präzisionsvorteile dieses Instrumentes, die neben der Erhebung eines über bis zu drei Jahren erstreckenden Kalendariums detaillierte Aussagen über bis zu vier Arbeitsstellen enthielten, hatten zur Folge, daß der sonst in der Methodenliteratur vorgeschlagene Umfang eines postalisch versendeten Fragebogens von 12 Seiten überschritten werden mußte und nunmehr 24 DIN A 5 Seiten betrug. Dadurch zu vermutende Reduzierungen der Rücklaufquoten wurden hier zugunsten des Ziels eines aussagekräftigen Verlaufsdatensatzes in Kauf genommen.

Die Entscheidung für den Weg der postalischen Erhebung ist organisatorisch, zeitlich und techisch mit einem enormen Aufwand verbunden. In jedem Fall ist der Erfolg der Erhebung sehr eng an die Frage der Erreichbarkeit der Untersuchungspersonen gekoppelt. Da die Erhebungen 1994 und 1997 in der Art der Durchführung variieren, ist eine getrennte Beschreibung der Feldphasen erforderlich.

2.4 Durchführung der ersten postalischen Befragung 1994

Eine Besonderheit der Basisbefragung der "Berufsverlaufsstudie Ostdeutschland" im Jahre 1994 war, daß der vorhandene Adressenbestand jeweils die Adresse eines Befragten zum Zeitpunkt des kohortendefinierenden Abschlusses im Jahr 1985 und 1990 enthielt. Bis zum Befragungszeitpunkt 1994/95 waren diese Adressen veraltet, da mit einer relativ großen Mobilität dieser jungen Befragtengruppe zu rechnen war. Zur Adressenaktualisierung wurden zwei Verfahren verwendet:

1.) Bei den Adressen der Hochschulabsolventen lagen zwei Adreßbestände vor, die aktuelle Adresse bei Studienabschluß (meist Studentenwohnheime), sowie die sogenannte "Heimatadresse", bei der es sich meist um die Anschrift der Eltern handelte. Bei den "Studentenadressen" war mit einem erheblichen Ausfall an Nicht-Erreichbarkeit zu rechnen, die in direktem Zusammenhang mit den Untersuchungszielen standen, da – eventuell aus beruflichen Gründen – mobilere Personen sehr viel weniger zu erreichen gewesen wären als weniger mobile. Bei den "Heimatadressen" war ebenfalls aufgrund der (lebenszyklisch geringeren) Mobilität von Eltern deren eventuellem Tod, Wohnungsauflösung durch Scheidung

[7] vgl. Anhang. Wir möchten Frau Prof. Dr. M. Buchmann für die Unterstützung bei der Konzeption des Fragebogens danken.

oder Abbruch der Beziehungen zwischen Eltern und Kindern mit Ausfällen zu rechnen. Bei einer Versendung der Fragebögen an die "Heimatadressen" war zudem mit einem gewissen "Filtereffekt" zu rechnen, da die Zielperson der Befragung nur erreicht werden konnte, wenn die Eltern des Befragten das Schreiben weiterleiteten. Insgesamt war in beiden Fällen mit erheblichen Ausfallraten aufgrund von Nicht-Erreichbarkeit der Befragten zu rechnen. Unser Projekt entschied sich vorwiegend aus methodischen Gründen für eine Verschickung über die "Heimatadressen", da die zu erwartenden Ausfälle bei dieser Verschickungsart im wesentlichen von Ereignissen bei den Eltern der Befragten abhängig waren. Die Verteilung dieser Ausfälle auf die Untersuchungspopulation kann methodisch als im wesentlichen "stichprobenneutral" angesehen werden, da Umzug und Tod der Eltern z.B. nicht in direktem Zusammenhang mit der Gestalt von Berufsverläufen von Hochschulabsolventen stehen. Bei einer Verschickung an die "Studentenadressen" wäre dagegen mit einem Ausfall zu rechnen gewesen, der in direktem Zusammenhang mit den Berufsverläufen der Absolventen stehen würde – der dadurch verursachte Ausfall hätte einen Stichprobenbias zugunsten von weniger mobilen Personen enthalten.[8]

2.) Die Bundespost hatte einen wesentlichen Anteil an der Adressenaktualisierung. Briefe, die an "Heimatadressen" verschickt wurden und mit einem entsprechenden Vermerk versehen waren ("Falls Empfänger verzogen, nachsenden!", "Anschriftenberichtigungskarte mit neuer Anschrift!", und "Falls unzustellbar, zurück"), konnten entweder von den "Heimatempfängern" (in der Regel Eltern) direkt ohne Kosten und ohne Öffnen des Briefes mit der Post weitergeleitet werden oder wurden mit neuer Adresse an die Versender zurückgeschickt. Dieses Verfahren hat im Unterschied zu einer Einholung von Adressenänderungen über die Meldeämter den Vorteil, daß es sehr viel kostengünstiger, schneller und weniger arbeitsintensiv abläuft. So waren die Anschriftenberichtigungskarten in der Regel innerhalb einer Woche wieder zurück und konnten zur Aktualisierung des Adressenstamms herangezogen werden. Die Post stellte dabei in ihrer Rolle als "Informationsvermittler" schneller als die Meldeämter die Verbindung zwischen elterlichen "Heimatadressen" und aktuellen neuen Adressen her. Dieses Verfahren stellte sich aufgrund der regen Beteiligung der Eltern als recht zuverlässig heraus.

Der Versand der Fragebögen erfolgte in drei Wellen. In der ersten Welle wurden die Fragebögen zusammen mit einem Anschreiben verschickt. In der zweiten Welle wurde ein Erinnerungsschreiben versendet und in der dritten Welle wurden noch einmal die gesamten Erhebungsmaterialien an die Befragten befördert. Der Abstand zwischen der ersten Versendung und der Verschickung eines Erinnerungsschreibens war abweichend von den Vorschlägen der Methodenliteratur mit vier Wochen relativ lang, da wir wußten, daß die von uns gewählte Verschickungsart relativ weite, d.h. zeitintensive Wege zur Folge hatte. Abb. 2.8 zeigt im Detail den Fragebogenrücklauf im Erhebungszeitraum.

[8] Bei dem in manchen Untersuchungen verwendeten Verfahren einer Adressenaktualisierung durch einen Aufruf in lokalen Zeitungen wäre ein entsprechender bias zugunsten Nicht-Mobiler zu erwarten gewesen.

Abb. 2.8.: Verschickungswellen und Rücklaufzahlen nach Gruppen

	Hochschulab-solventen 1985	Hochschulab-solventen 1990	Abs. einer be-ruflichen Lehre 1985	Abs. einer be-ruflichen Lehre 1990	Gesamt
1. Fragebogenversand (19.9.94)	2315	2223	2000	1998	8536
Rücklauf bis 14.10.94	269	362	114	169	914
Verschickung des Erin-nerungsschreibens (14.10.94)	5802				
Rücklauf bis 25.10.94	299	387	124	183	993
2. Fragebogenversand (4.11.94)	4884				
Rücklauf bis 23.1.95	551	666	408	505	2130

Es zeigt sich, daß die "langen" Wege der Verschickung nach der ersten Verschickung zu einer relativ geringen Rücklaufquote führten. Das Erinnerungsschreiben führte ebenfalls nicht zu dem erwünschten Effekt einer bedeutenden Rücklaufsteigerung. Erst nach der zweiten Verschickung des Fragebogens erhöhte sich der Rücklauf deutlich. Bei der Verschickung des Erinnerungsschreibens und bei der zweiten Verschickung des Fragebogens konnte die Zahl der Sendungen deutlich reduziert werden, da bereits zu diesem Zeitpunkt von vielen Befragten bekannt war, daß ihre Adresse auf dem gewählten Verfahren nicht eruierbar war. Bei den Angaben des Rücklaufs zum 23.1.95 handelt es sich bereits um "bereinigte Daten", bei denen nur Teilnehmer der Befragung erfaßt wurden, die den Kriterien der Untersuchung (Kohorte, Ausbildungsabschluß, Ausbildungsort) genügten. Die folgenden Abbildungen geben einen genauen Aufschluß über den Verbleib der verschickten Fragebögen.

Abb. 2.9: Verbleib der Fragebögen nach Bruttosample

	Universität 1985		Universität 1990		Berufliche Lehre 1985		Berufliche Lehre 1990		
	n	%	n	%	n	%	n	%	Gesamt
Bruttosample	2315	100	2223	100	2000	100	1998	100	8.536
nicht erreichbar (unbe-kannt verzogen)	891	39	643	30	576	29	423	21	2.533
zurückgeschickte Frage-bögen	658	29	773	36	476	24	572	29	2.479
Davon:									
- Teilnahme	551	24	666	31	408	21	505	25	2130
- explizite Verweige-rung	36	1	27	1	44	2	26	1	133
- falsche Kohorte/Ab bruch der Ausbildung	71	3	80	4	24	1	41	2	216
keine Antwort	766	33	807	36	948	47	1003	50	3524

Abb. 2.10.: Verbleib der Fragebögen nach Nettosample 1

	Universität 1985		Universität 1990		Berufliche Lehre 1985		Berufliche Lehre 1990		
	n	%	n	%	n	%	n	%	Gesamt
Nettosample	1424	100	1580	100	1424	100	1575	100	6003
zurückgeschickte Fragebögen	658	47	773	49	476	34	572	36	2.479
Davon:									
- Teilnahme	551	39	666	45	408	29	505	32	2130
- explizite Verweigerung	36	3	27	2	44	3	26	2	133
- falsche Kohorte/Abbruch der Ausbildung	71	5	80	5	24	2	41	3	216
keine Antwort	766	53	807	51	948	66	1002	64	3524

Die Abbildung 2.10 zeigt, daß - obwohl die Adreßbestände 5 bzw. 10 Jahre alt waren - durch den gewählten Verschickungsweg ca. zwei Drittel der Personen des Bruttosamples noch erreicht werden konnten. Je älter die Adressen waren, desto geringer ist diese Zahl. Es ist zu berücksichtigen, daß bei der Verschickung der Fragebögen bei den Absolventen der Universität Rostock bekannt war, daß sich in unserem Adreßbestand 694 Studienabbrecher bzw. Studienwechsler befanden, die in den Jahren 1985 und 1990 die Universität Rostock ohne einen Studienabschluß verlassen hatten. Mit den vorhandenen Daten konnte diese nicht zu unserer Grundgesamtheit gehörige Personengruppe *vor* der Fragebogenverschickung nicht aussortiert werden (vgl. Abb. 2.9.). Nur ein Teil dieser Personen hat geantwortet und konnte nach der Rücksendung aus dem Datensatz entfernt werden. Für das Netto-Sample wurde die Gruppe der Studienabbrecher und Studienwechsler aus der Betrachtung ausgeschlossen, da für diese Gruppe mit der Beantwortung der ersten Frage des Fragebogens klar sein mußte, daß sie nicht Teil der Untersuchungsgruppe war. Diese Gruppe konnte sowohl Teil der unbekannt verzogenen Personen als auch der Gruppe der nicht Antwortenden sein. Um diese Gruppe nicht zu überschätzen, wurde bei der Bereinigung des Netto-Samples davon ausgegangen, daß sie mit gleicher Wahrscheinlichkeit in der Gruppe der unbekannt Verzogenen vertreten war (Abb. 2.11.).[9]

[9] Um die Schätzung nachvollziehbar zu machen, wurde in Abbildung 2.11 jeweils die Berechnungsweise der Beschreibung der Kategorie hinzugefügt. Bei Zeile 8 lautet die Formel z.B. 6/1 * 3, was bedeutet, daß der Wert von Zeile 6 durch den Wert von Zeile 1 geteilt wurde und dann mit dem Wert aus Zeile 3 multipliziert wurde.

Abb. 2.11.: Bereinigung des Nettosamples

	Hochschulabsolventen 1985		Hochschulabsolventen 1990	
	n	%	n	%
1. Bruttosample	2315	100	2223	100
2. zurückgeschickte Fragebögen	658	29	773	36
3. nicht erreichbar (unbekannt verzogen)	891	39	643	30
4. keine Antwort	766	33	807	36,3
5. Nettosample (Bruttosample - Nicht-Erreichbare 1-3)	1424	100	1580	100
6. Anzahl der Studienabbrecher / Studienwechsler im Bruttosample	353		341	
7. davon geantwortet	71		80	
8. davon "nicht erreichbar" (Schätzung: 6/1*3)	136		103	
9. davon "keine Antwort" (Schätzung: 6/1*4)	117		115	
10. korrigiertes Nettosample (Bruttosample - (Nicht-Erreichbare + Studienabbrecher/Studienwechsler) [5-7-9])	1236		1385	

Abb. 2.12.: Rücklaufquoten bezogen auf das Nettosample

	Nettosample (Bruttosample abzgl. Nicht-Erreichbare / Ausbildungsabbrecher / falsche Kohorte)	verwertbare Rückläufe	Responseraten
Hochschulabsolventen 1985	1236	551	44,6 %
Hochschulabsolventen 1990	1385	666	48,1 %
gesamt	2621	1217	46,4 %
Berufliche Lehre 1985	1398	408	29,2 %
Berufliche Lehre 1990	1532	505	33,0 %
gesamt	2930	913	31,2 %
Insgesamt	5551	2130	38,4 %

In Abb. 2.12. wird die Rücklaufrate in Relation zu den erreichten Personen der Grundgesamtheit angegeben.

In Kapitel 4 werden für die einzelnen Fächergruppen die nach Abschlußjahrgang und Abschlußort differenzierten Rücklaufquoten mit den Verteilungen der Grundgesamtheit abgebildet.

2.5 Durchführung der zweiten postalischen Befragung 1997

Eine Besonderheit der zweiten Befragung lag darin, daß der vorhandene Adressenbestand gegenüber der Datenbasis 1994 relativ aktuell war. Dies betrifft insbesondere das Adressenmaterial der Teilnehmer aus der Basiserhebung. So wurden die Befragten der Basiserhebung im Rahmen der Panelpflege von 1994 bis 1996 einmal jährlich angeschrieben und um Aktualisierung ihrer Anschriften gebeten. Darüber hinaus wurden kontinuierlich mit jeder neuen Verschickungswelle die Anschriften aller auf dem Postwege nicht erreichbaren Personen über die Einwohnermeldeämter in Erfahrung gebracht. Von den 1667 Personen der Basiserhebung 1994, die sich 1996 bereit erklärten, sich 1997 gegebenenfalls erneut befragen zu lassen, haben somit immerhin 87,8% tatsächlich an der Befragung teilgenommen (vgl. Abbildung 4.8.) und nur 17 Personen (1 %) konnten nicht ausfindig gemacht werden. Der Rücklauf war für diese Personengruppe somit gemessen am Adressenbestand hoch, und technisch gesehen kann angesichts der geringen 'Schwundmenge' von einem Prozent von einem durchaus positiven Verfahrensweg gesprochen werden. Der hohe Rücklauf ist nicht zuletzt einer intensiven Panelpflege zu verdanken.

Eine weitere Besonderheit der zweiten Befragung war, daß es sich bei den Anschriften der Absolventen des Jahres 1995 im Gegensatz zu der Basisererhebung nicht um reine 'Elternanschriften' handelte, sondern um die aktuelle Wohnanschrift zum Zeitpunkt des Ausbildungsabschlusses. Dieser Adressenbestand war zwar gegenüber der Datenbasis 1994 immer noch relativ aktuell, dennoch ist nach rund zwei Jahren mit einem Ausfall in der Höhe von 10 bis 20 Prozent zu rechnen gewesen, der in direktem Zusammenhang mit den Berufsverläufen der Absolventen stehen würde, d.h. weniger mobile Personen wären eher zu erreichen gewesen als mobile Personen. Gleichfalls war zu vermuten, daß dieser Ausfall ohne 'Korrekturleistungen' zu einem Stichprobenbias zugunsten von weniger mobilen Personen geführt hätte. Um diesen Stichprobenbias so gering wie möglich zu halten, folgte man der Gruppe der hochmobilen Absolventen bis zum ersten, zweiten und dritten Umzug. Wie schon in der Basiserhebung hatte die Bundespost einen wesentlichen Anteil an der Adressenaktualisierung. Sie stellte in 84% aller Fälle schnell und zuverlässig die Verbindung zwischen der alten Heimatanschrift und der aktuellen neuen Adresse her. Anschriften, die über den postalischen Weg nicht erreicht werden konnten, wurden über die Meldeämter aktualisiert. Insgesamt konnten 905 Personen über die Postadresse nicht erreicht werden, d.h. 15,97 % aller Anschriften waren zum Zeitpunkt der ersten Verschickung nicht mehr aktuell. Im Gegensatz zur Basiserhebung wurden über diese Personen bei den zuständigen Einwohnermeldeämtern Auskünfte über die aktuelle Wohnanschrift eingeholt.

Der gewählte Weg über die Einwohnermeldeämter sah folgendermaßen aus: Prinzipiell besteht in der Bundesrepublik Deutschland eine gesetzliche Meldepflicht, d.h. alle Bürger müssen sich bei den für sie zuständigen Meldebehörden unter ihrer Wohnanschrift anmelden und gegebenenfalls unter Angabe der neuen Wohnanschrift abmelden. Es kann daher mit einer großen Wahrscheinlichkeit davon ausgegangen werden, daß die 'unerreichbaren' Person unter der alten Anschrift polizeilich gemeldet waren und sich nach einem Umzug unter Angabe der neuen Wohnanschrift abgemeldet haben. In dem Fall der 'Unerreichbarkeit' wurde das zuständige Einwohnermeldeamt angeschrieben und um eine einfache 'Meldeauskunft', d.h. die Angabe der aktuellen Wohnanschrift, gebeten. Diese einfache 'Meldeauskunft' war

i.d.R. kostenpflichtig und zog über die Grundgebühr hinausgehend noch weitere Folgekosten (z.B. Portokosten) nach sich. So gehörte es vielerorts zum allgemeinen verwaltungstechnischen Prozedere, die anfallenden Gebühren über den postalischen Weg eines 'Nachnahmeschreibens' einzuziehen bzw. die Gebühren vorab an die Meldestelle zu überweisen. Damit fielen nicht selten schon im Vorfeld der Erhebung Kosten in Höhe von 15-20 DM pro Einwohnermeldeamtziehung und Person an. Interessanterweise verzichteten auch viele Meldestellen auf die Gebühren zugunsten des Forschungsprojektes. Ein diesbezügliches Anschreiben hat uns vielerorts die Türen geöffnet. Das Antwort*verhalten* der kontaktierten Meldebehörden ist in jedem Fall positiv zu bewerten: so lagen am Ende der Erhebungsphase im Dezember 1997 lediglich über 14 Personen keine Auskünfte vor. Der *Nutzen* dieser Auskünfte war dagegen nicht immer sehr hoch. So kann die Antwort einer Einwohnermeldeamtziehung sowohl negativ als auch positiv ausfallen, d.h. bei einer negativen Antwort ist die Person definitiv für immer verloren und es gibt keinen Weg sie jemals wieder ausfindig zu machen. Und eine positive Antwort bedeutet wiederum nicht, daß die besagte Person tatsächlich auch unter der neuen Anschrift zu erreichen ist. Generell gibt es drei Möglichkeiten: 1. Es handelt sich bei der mitgeteilten neuen Anschrift tatsächlich um die aktuelle Anschrift. Das gewünschte Ziel wurde erreicht und der Fragebogen kann zugestellt werden. 2. Es handelt sich bei der mitgeteilten neuen Anschrift um die falsche Person. Nur wenn sich diese Person zurückmeldet, wissen wir tatsächlich, daß wir die falsche Person angeschrieben haben. In beiden Fällen (Rückmeldung und keine Rückmeldung) ist die ursprüngliche Person ebenfalls für immer verloren, da das konsultierte Einwohnermeldeamt mehrere Personen gleichen Namens im Register hat. 3. Es handelt sich bei der mitgeteilten neuen Anschrift nicht um die aktuelle Anschrift. Es ist davon auszugehen, daß sich die besagte Person ordentlich in ihrem Ursprungsort abgemeldet hat. Allerdings ist sie nach dem ersten Umzug ein weiteres Mal umgezogen. In diesem Fall muß an das zuständige zweite und wenn erforderlich auch dritte Einwohnermeldeamt herangetreten werden. Dieses zeit- und auch kostenintensive Verfahren hat sich über neun Monate hingezogen. So lagen bspw. erst nach drei Monaten (26.06.1997) von 905 Einwohnermeldeamtsanfragen 579 positive Einwohnermeldeamtsauskünfte (64% aller Personen) vor. Insgesamt jedoch konnten nach Abschluß des Verfahrens lediglich 158 Personen (17,5% von 905 Personen) nicht über das Einwohnermelderegister ausfindig gemacht werden und über 14 Personen (1,5% von 905 Personen) wurde von den angeschriebenen Meldebehörden keine Auskünfte erteilt. Das Verfahren wurde im Dezember 1997 abgeschlossen. Mit diesem technisch gesehen relativ aufwendigen Verfahren gelang es auch noch Personen nach dem ersten, zweiten oder dritten Umzug ausfindig zu machen, womit zumindest der hochmobilen Gruppe unter den Absolventen Rechnung getragen wurde.

Der Versand der Fragebögen erfolgte, wie auch bei der Basiserhebung, in mehreren Wellen. In der ersten Welle wurden die Fragebögen zusammen mit einem Anschreiben verschickt. In der zweiten Welle wurde ein Erinnerungsschreiben versendet und in der dritten Welle wurden noch einmal die gesamten Erhebungsmaterialien an die Befragten befördert. Die Befragten der ersten Welle wurden darüber hinaus noch ein viertes Mal mit einem Fragebogen angeschrieben. Der Abstand zwischen der ersten Versendung und der Verschickung eines Erinnerungsschreibens war aufgrund der zeitintensiven Wege wiederum relativ lang.

Durchlauf ohne Einwohnermeldeamtziehung:

		Anzahl	*Unzustellbarkeiten*	*Rücklauf bis zum Termin*
1. Fragebogenversand I	**27.03.1997**	5667	905	
2. Erinnerung	**24.04.1997**	3700		1062
3. Fragebogenversand II	**13.05.1997**	2907		1855
4. Fragebogenversand III	**26.06.1997**	255		

Der obigen Abbildung kann man entnehmen, daß die "langen" Wege der Verschickung zu einem langsamen Anstieg der Rücklaufquote führten. So waren bis zum Verschicken des ersten Erinnerungsschreibens von 5667 angeschriebenen Personen erst 18,74 % aller Fragebögen zurückgekommen und rund 16% aller Erstanschriften erwiesen sich als unbrauchbar und mußten über die Einwohnermeldeämter aktualisiert werden. Das Erinnerungsschreiben führte zu einem Anstieg der Rücklaufquote auf 32,8%. Erst nach der zweiten und dritten Verschickung der Fragebögen und der Einwohnermeldeamtsziehungen erhöhte sich der Rücklauf deutlich auf rund 56%. Die folgenden Abbildungen geben einen genauen Aufschluß über den Verbleib der verschickten Fragebögen.

Abb. 2.13.: Verbleib der Fragebögen nach Bruttosample

Befragung 1997 (2. Panel)	**Absolventen 1985 und 1990**		**Absolventen 1995**		
	n	%	n	%	Gesamt
Gesamtsample (Bruttosample)	1.667	100	4.000	100	5.667
nicht erreichbar (unbekannt verzogen)	17	1	141	3,5	158
zurückgeschickte Fragebögen	1468	88,1	1701	42,5	3169
Davon:					
- Teilnahme	1463	87,8	1646	41,1	3109
- darunter anonym	-	-	64	3,9	
- explizite Verweigerung	5	0,3	31	0,8	36
- falsche Kohorte/Abbruch der Ausbildung		-	24	0,6	24
keine Antwort	182	10,9	2158	53,9	5667

Die Abbildung 2.13. zeigt, daß mit dem gewählten Verschickungsweg über 97,2 % der Personen des Gesamtsamples erreicht werden konnten, d.h. nur 158 Personen oder 2,8% aller Personen konnten über die Einwohnermeldeämter nicht mehr ermittelt werden. Den höchsten Rücklauf weisen die Teilnehmer der 1994 erstmals befragten Personen mit 87,8% auf. Die hohe Quote unterstreicht die Wichtigkeit der Bedeutung von aktiven Maßnahmen der Panelpflege. Im Rahmen der Berufsverlaufsstudie Ostdeutschland wurde eine aktive Panelpflege betrieben, indem bspw. einmal im Jahr eine Sonderbriefmarke und ein Dankschreiben an alle Teilnehmer verschickt wurde. Darüber hinaus wurde am Ende des Projekts an alle Lehrabsolventen ein kleines Dankeschön im Werte von DM 11,- in Briefmarken und an alle Teilnehmer eine Informationsbroschüre verschickt. Von den Absolventen des Jahres 1995 liegen rund 41% brauchbare Fragebögen vor. Auch hier erhielten alle Lehrabsolventen ein kleines Dankeschön im Werte

von DM 11,- in Briefmarken und eine Informationsbroschüre. Nach Bereinigung des Gesamtsamples um die Zahl der 'Unzustellbarkeiten' ergibt sich eine Gesamtquote von 56,7%.
Es ist zu berücksichtigen, daß bei der Verschickung der Fragebögen bei den Adressen der Absolventen einer beruflichen Lehre bekannt war, daß sich in unserem Adreßbestand auch Absolventen befanden, die im ersten Anlauf ihrer beruflichen Abschlußprüfung (meist IHK) scheiterten und diese Prüfung nach einigen Monaten wiederholten. Diese Absolventen wurden von den berufsbildenden Schulen mit dem Ende der Berufsschulpflicht als Abgänger geführt, obwohl die wiederholende Abschlußprüfung bspw. sechs Monate später war. Mit den vorhandenen Daten konnte diese nicht zu unserer Grundgesamtheit gehörige Personengruppe vor der Fragebogenverschickung nicht aussortiert werden. Nur ein Teil dieser Personen (24) hat geantwortet und konnte nach der Rücksendung aus dem Datensatz entfernt werden. Diese Gruppe wurde aus der Betrachtung ausgeschlossen, da für sie mit der Beantwortung der ersten Frage des Fragebogens klar sein mußte, daß sie nicht Teil der Untersuchungspopulation war. Aufgrund der geringen Zahl wurde keine Korrektur des Samples vorgenommen.

2.6 Bewertung der Verschickungsaktion

Zusammenfassend ist der verwertbare Rücklauf von 2130 Fragebögen nach der ersten Befragungswelle und 3109 nach der zweiten Befragungswelle für das Projekt sehr erfolgreich. Mit diesen hohen Rückläufen wurden auch in kleineren Untergruppen noch große Fallzahlen erreicht, die vielfältige Verlaufsdatenberechnungen zulassen. Die Verteilungen nach Ausbildungsort und –jahr, Abschlußart und Geschlecht (Abb. 3.1) belegen dieses recht anschaulich. Damit ist eine gute Grundlage für eine eventuelle Wiederholungsbefragung im Jahre 2000 gelegt worden. Bei einer Beurteilung der Rücklaufraten ist auch zu beachten, daß es sich um eine schriftliche Befragung mit einem relativ umfangreichen Fragebogen (30 bzw. 24 Seiten) und einem anspruchsvollen Gegenstand (detaillierte zeitliche Darstellung eines Erwerbsverlaufs über 5 bzw. 10 Jahre) handelt. Die Rücklaufquote bei den Absolventen einer beruflichen Lehre fällt im Vergleich dazu geringer aus. Über die Gründe kann hier nur spekuliert werden. Aus der Methodenliteratur ist bekannt, daß Rückläufe in statusniedrigeren Gruppen geringer sind. Im speziellen Fall unserer Untersuchung kommt hinzu, daß diese Gruppe stärker von negativen Folgen des Transformationsprozesses betroffen ist. Daß dies Folgen für die Antwortbereitschaft hat, zeigte sich schon in der qualitativen Phase unseres Projektes, bei der die Teilnahmebereitschaft von Absolventen einer beruflichen Lehre deutlich geringer war als die Teilnahmebereitschaft von Hochschulabsolventen. Rücklaufsenkend wirkte sich unter Umständen auch aus, daß diese Gruppe aufgrund des niedrigeren Bildungsniveaus mehr Schwierigkeiten mit dem Ausfüllen des Fragebogens hatte.

3 Die Datenbank der Berufsverlaufsstudie Ostdeutschland

3.1 Das Nettosample nach der zweiten Erhebungswelle

Das strukturierende Element der Datenbasis der 'Berufsverlaufsstudie Ostdeutschland' ist das Kohortendesign, welches sich in besonderer Weise zur Erforschung des **sozialstrukturellen Wandels in Ostdeutschland** eignet. Die drei Absolventenkohorten verweisen nämlich auf drei unterschiedliche Ausgangssituationen zum Zeitpunkt des Berufseinstiegs:

1. Die Absolventen des Jahres 1985 haben die ihre berufliche Ausbildung und den Berufseinstieg noch unter DDR-Bedingungen vollzogen.
2. Die Absolventen des Jahres 1990 haben ihre berufliche Ausbildung zwar noch unter DDR-Bedingungen vollzogen, traten aber zum Zeitpunkt der „Wende" in den Arbeitsmarkt ein.
3. Die Absolventen des Jahres 1995 haben sowohl ihre berufliche Ausbildung als auch den Berufseinstieg nach der Maueröffnung vollzogen.

In der Abbildung 3.1. sind die Fallzahlen aller drei Abschlußjahrgänge differenziert nach der Abschlußart, dem Geschlecht und dem Ausbildungsort dargestellt. 1994 wurden die Kohorten 1985 und 1990 erstmals befragt. Damals haben insgesamt 959 Personen des Jahres 1985 und 1.171 Personen des Jahres 1990 teilgenommen. 1997 wurde eine weitere Kohorte hinzugenommen: Es haben 1.646 Absolventen des Jahres 1995 teilgenommen.

Abb. 3.1: Grundauszählung nach Abschlußjahrgang, jeweils die ersten Erhebungen

	Abschlußjahr 1985		**Abschlußjahr 1990**		**Abschlußjahr 1995**		**Zeilen gesamt**	
	n	%	n	%	n	%	n	%
Abschlußart								
Lehre	408	42,5	505	43,1	591	35,9	1.504	39,8
Studium	551	57,5	666	56,9	1055	64,1	2.272	60,2
Spalte gesamt	959	100,0	1.171	100,0	1.646	100,0	3.776	100
Geschlecht								
Männer	461	48,1	513	43,8	805	48,9	1.779	47,1
Frauen	498	51,9	658	56,2	840	51,1	1.997	52,9
Spalte gesamt	959	100,0	1.171	100,0	1.646	100,0	3.776	100
Ausbildungsort								
Rostock	475	49,5	535	45,7	812	49,3	1.822	48,2
Leipzig	429	44,7	584	49,9	829	50,7	1.842	48,8
nicht zuordbar	55	5,8	52	4,4	-	-	112	3,0
Spalte gesamt	959	100,0	1.171	100,0	1.646	100,0	3.776	100,0

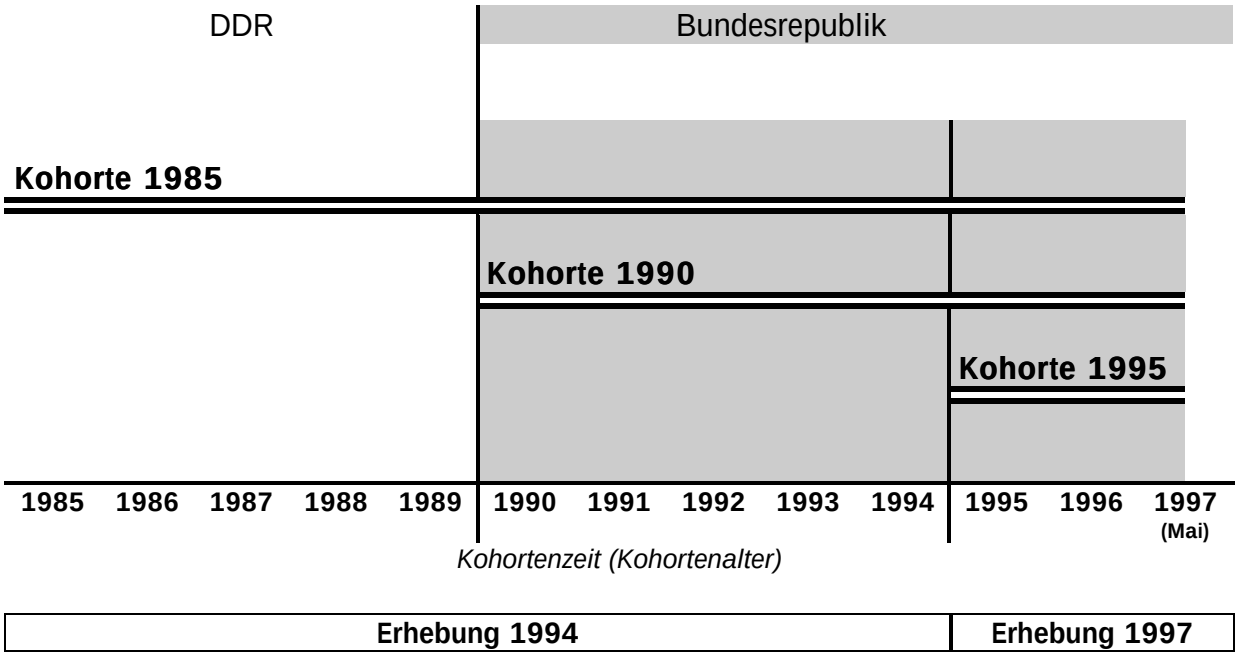
Bemerkenswert ist die Kohortenstärke aller drei Jahrgänge: So variiert die Kohortenstärke erheblich zwischen den einzelnen Abschlußjahrgängen. Der Anteil der 85er Kohorte am Gesamtsample beträgt nunmehr 25,4%, der Anteil der 90er Kohorte 31,0% und der Anteil der 95er Kohorte 43,6%. Die Teil-

nahmebereitschaft hängt demnach in erster Linie vom *Alter* der Absolventen ab, denn zum Zeitpunkt der Erstbefragung waren die Kohorten 1985 und 1990 im Durchschnitt älter als die Kohorte 1995. Hinzu kommt aber sicherlich auch der Einfluß des unterschiedlichen Anschreibeverfahrens. So sind wir 1994 an die Absolventen der Jahrgänge 1985 (Bruttosample ca. 4000) und 1990 (Bruttosample ca. 4000) über die Elternanschriften herangetreten. 1997 haben wir dagegen den direkten Weg über die zuletzt gültige Anschrift (Bruttosample 4000) gewählt.

Über die Kohortenfrage hinaus bieten sich aufgrund der hohen Fallzahlen folgende Differenzierungslinien an: Geschlecht, Bildung (Lehre/Studium), Region und beispielsweise Berufsgruppen. Die hohen Fallzahlen erlauben selbst bei den Berufsgruppen einen Kohortenvergleich. Bei Kohortenvergleichen ist jedoch zu beachten, daß datentechnisch ein sehr eng am Ausbildungsabschluß orientiertes Beobachtungsfenster vorliegt. In Abhängigkeit von der Kohortenlage liegen nach der zweiten Befragungswelle 1997 Angaben über einen Zeitraum von maximal 12 (85er Kohorte), 7 (90er Kohorte) und 2 (95er Kohorte) Jahren vor.

Die mit der jeweils ersten Befragung erfaßten Zeiträume sind der Abbildung 3.2. zu entnehmen.

Abb. 3.2. Kohortenlage (Kohortenalter) bei der Berufsverlaufsstudie Ostdeutschland



3.2 Panelmortalität

Ziel der zweiten Erhebung im Jahre 1997 war neben der Einbindung einer weiteren Kohorte die Verlängerung des Beobachtungszeitraumes (Zeitfenster) für die 85er und 90er Kohorte. Über die Verlängerung des Beobachtungszeitraumes versprach man sich langfristige Prozesse der De- und Restabilisierung beobachten zu können. Mit Hilfe der nächsten Abbildung kann man sich einen Eindruck über das Teilnehmerverhalten der Absolventen der Kohorten 1985 und 1990 verschaffen.

Abb. 3.3.: Panelmortalität: II. Welle (Kohorten 1985 und 1990)

	Abschlußjahr 1985 I+II			Abschlußjahr 1990 I+II		
	Welle I (absolut)	Welle II (absolut)	Verlust in %	Welle I (absolut)	Welle II (absolut)	Verlust in %
Abschlußart						
Lehre	408	247	-39,5	505	300	-40,6
Studium	551	422	-23,4	666	494	-25,8
Spalte gesamt	959	669	-30,2	1.171	794	-32,2

Von den 959 Befragten der 85er Kohorte waren bis 1997 real 669 Personen zu einem weiteren Interview bereit. Der Verlust beträgt 30,2 Prozent. Bei der 90er Kohorte ist der Verlust sogar etwas größer: Von den 1.171 Personen haben 1997 nur 794 geantwortet, was einem Verlust von 32,2 Prozent entspricht. Insgesamt ist die Zahl der Personen von 2.130 auf 1.463 gesunken, die Gesamtverlustquote beträgt damit 31,3 Prozent. Hierbei ist jedoch zu bedenken, daß der hohe Schwund zum Teil auch auf die 1996 durchgeführte Datenschutzauflage (Einverständniserklärung für die Befragung 1997) zurückzuführen ist, d.h. von den 31,3 Prozent entfallen direkt 22 Prozent auf die Datenschutzauflage und knapp 10 Prozent auf die zweite Befragung.

Bemerkenswert sind hier leider auch die Verschiebungen innerhalb der Abschlußart: Betrug der Anteil der Lehrabsolventen 1994 noch 42,9 %, ist er nunmehr mit der zweiten Erhebungswelle auf 37,4 Prozent gesunken. Bei beiden Kohorten beträgt die Verlustquote der Lehrabsolventen ungefähr 40%, bei den Akademikern liegt sie bei rund 26%. Für tiefergehende Verlaufsdatenberechnungen des Periodenzeitraumes 1990 bis 1997 könnte es hier zu Verzerrungen aufgrund zu geringer Fallzahlen kommen.

In Abb. 3.4 ist zu sehen, daß insbesondere der lange Zeitraum zwischen der Erhebung 1994 und der datenschutzrechtlich erforderlichen Einverständniserklärung der Adreßaufbewahrung 1996 zu einer erhöhten Panelmortalität führte. Bei den Einverständniserklärungen für die Adreßaufbewahrung für die Erhebung im Jahre 2000 konnte durch einen kürzeren Zeitabstand die Panelmortalität auf diesem Wege gesenkt werden. Zur Zeit haben von 3.030 angeschriebenen Personen 2.571 ihre Einverständniserklärung für das Jahr 2000 gegeben und 107 haben explizit abgelehnt (Abb. 3.4.).

Abb. 3.4.: Panelmortalität nach Einverständniserklärung für die dritte Welle (Stand 02.6.98)

	Abschlußjahr 1985 I+II				Abschlußjahr 1990 I+II			
	Welle I (absolut)	Welle II (absolut)	Welle III	Verlust in %	Welle I (absolut)	Welle II (absolut)	Welle III	Verlust in %
Abschlußart								
Lehre	408	247	223	-9,71	505	300	262	-12,66
Studium	551	422	383	-9,24	666	494	448	-9,31
Spalte gesamt	959	669	606	-9,41	1.171	794	710	-10,57

Bei Kohortenvergleichen ist auch hier zu beachten, daß datentechnisch nur ein sehr eng am Ausbildungsabschluß orientiertes Beobachtungsfenster vorliegt. Von 2.130 1994 erstmals befragten Personen

konnten immerhin noch 1.463 Personen ein zweites Mal befragt werden. 667 Personen oder 31,3% der Ursprungspopulation standen für eine weitere Befragung nicht zur Verfügung. Für 669 Absolventen des Jahres 1985 liegen nun Angaben über einen Zeitraum von maximal 12 Jahren vor (85er Kohorte). Für 794 Absolventen des Jahres 1990 liegen Angaben über einen Zeitraum von maximal 7 Jahren vor.

3.3 Datenedition

Bei der Datenerfassung, -prüfung und -bereinigung handelt es sich um einen mehrstufigen Prozeß mit aufeinander aufbauenden Arbeitsschritten:

1. Erfassung und Prüfung des Rücklaufs anhand der Adressenprotokolle
2. Fragebogenprüfung
3. Datenerfassung und Übertrag auf 'Code-Sheets'
4. Dateneingabe
5. Automatische Datenprüfung/Kontrollroutinen

Die erste Sichtung der Fragebögen sowie die Vorkontrolle erfolgten direkt mit Posteingang. Nach der Registratur und ersten Kontrolle wurden die Fragebögen zur weiteren Erfassung der Rohdaten einem speziell geschulten Stab von studentischen Hilfskräften übergeben. Zur Vermeidung von Erhebungsfehlern (z.B. die Zeitstruktur von Längsschnittdaten verbietet bei der Ereignisanalyse chronologische 'Überlappungen') waren hierbei mehrere Schritte vorgeschaltet:

1. Es wurde ein Team von zehn studentischen Hilfskräften mit den Besonderheiten der vorliegenden Zeitdaten, den einzelnen Variablen und ihren Merkmalsausprägungen und den Datenbankstrukturen ausführlich vertraut gemacht. Hierzu erfolgten sowohl Einzel- als auch Gruppenschulungen und darüber hinaus wurden täglich Codiersitzungen zur Klärung von Einzelfragen einberufen. Alle Entscheidungen zur Codierung wurden protokolliert und sofort in Textform dem Codebuch beigeheftet. Die Dateneingeber wurden ebenfalls nach einer programmtechnischen Einweisung mit der besonderen Datenstruktur und möglichen Problemen vertraut gemacht.
2. Da sich in der quantitativen Sozialforschung das Übertragen der Rohdaten auf sogenannte 'Codesheets' bewährt hat, erfolgte die Erfassung der Rohdaten durch den Übertrag auf eigens vorbereitete Vordruckformulare. Für jedes der vierzehn Unterfiles stand ein eigenes Erhebungsblatt zur Verfügung, das die Besonderheiten und Strukturen der jeweiligen Datei berücksichtigte und effektiv die Dateneingabe unterstützte. Eine Direkteingabe vom Fragebogen in den Computer hätte erstens eine spätere Kontrolle der Daten wesentlich erschwert, zweitens bei inhaltlichen Unstimmigkeiten (Fehlangaben der Befragten) Einzelentscheidungen in den alleinigen Ermessensspielraum der Eingeber überantwortet, drittens die Sichtung der Fragebögen auf den Zeitpunkt der Eingabe hinausgezögert und den Eingabeprozeß durch den Zusammenfall aller Arbeitsschritte immens verlängert. Umfang und Komplexität des Fragebogens, die Struktur der Datenbank und die Produktivität im Eingabeprozeß sprachen gegen eine Direkteingabe. Durch den Übertrag der Daten auf die 'Codesheets' waren die Erfasser gezwungen, sich mit der komplexen Biographie der Teilnehmer inhaltlich auseinanderzusetzen und Unstimmigkeiten in den Codiersitzungen gemeinsam zu klären.

In einem dritten Schritt wurden die Daten räumlich getrennt von den Erhebern von zwei eigens für die Dateneingabe angelernten studentischen Hilfskräften in den Computer eingegeben. Die ersten Kontrollroutinen und Testläufe unterlagen ihrem Verantwortungsbereich. Da die Daten in den 'Codesheets' für jedes Unterfile untereinander standen, konnten an dieser Stelle sehr schnell weitere Ungereimtheiten aufgedeckt und geklärt werden. Ferner war die Eingabemaske so programmiert, daß für jede Variable die Eingabebandbreite auf Zahl ihrer Merkmalsausprägungen festgelegt war. Die Maske ließ es nicht zu, daß ein Absolvent der 95er Abschlußkohorte als Abschlußdatum z.B. 1994 eingegeben bekam oder ein männlicher Teilnehmer eine Mutterschaft im Kalendarium angab. Derartige Logik-Brüche konnten somit schon über die Maskenprogrammierung erkannt und korrigiert werden. In einem letzten Schritt wurden Stichprobenuntersuchungen vorgenommen, Kontroll- und Plausibilitätsroutinen durchgeführt und Korrekturen vorgenommen.

3.4 Datenaufbau und Organisation

Vor dem Verschicken der Fragebögen wurde aufbauend auf dem Codebuch der ersten Erhebung ein ausdifferenziertes Codebuch für den späteren Datenbank-Aufbau entwickelt, in dem für jede der Variablen der Kurzname, das Label, der Charakter (numerisch oder alphanumerisch), die Zahl der Zeichen und der spätere Speicherort verbindlich festgelegt wurde. Aus Gründen des besseren Datenmanagements wurde beschlossen, die Daten getrennt nach ihren thematischen Schwerpunkten (räumliche Mobilität, Schulbildung, Weiterbildung, Erwerbskalendarium, Standarddemographie etc.) in insgesamt vierzehn Datensätzen zu organisieren. So wurden in der ersten Datei z.B. die zeitunabhängigen standarddemographischen Angaben zum Geschlecht, Alter etc. und in der zweiten Datei alle Querschnittangaben zu aktuellen Einstellungsfragen abgespeichert. In den verbleibenden Dateien wurden die zeitbezogenen Daten der Befragten in chronologischer Reihenfolge abgespeichert. Die Datenbankstruktur erlaubt in Abhängigkeit von der jeweiligen Fragestellung möglichst einfache Datenretrievals. Durch die Verhinderung von Redundanzen kommt es zu einer minimalen Speicherbelegung.

Die Erhebung von Massendaten bringt zweifelsohne immer auch die Frage nach dem besten oder angemessensten Programm auf. Neben reinen Datenbankverwaltungsprogrammen (wie Oracle, SIR, DBASE) bieten sich eine Fülle von recht einfachen bis hochkomplizierten Programmen an, die alle für die Organisation und Verwaltung von Massendaten geeignet wären. Da aber hierbei der Aufwand durch Einarbeitung, Schulung und Hardwareeinsatz oftmals eher kontraproduktive Formen annimmt, einigten wir uns nach Absprache mit der Methodenabteilung und erfahrenen Datenspezialisten auf das SPSS-Unterprogramm DATA ENTRY. Dieses Programm ist für die effektive und rationale Eingabe von Massendaten sehr gut geeignet, ist programmtechnisch sehr sicher (keine Systemabstürze) und zieht einen minimalen Aufwand an Programmschulung nach sich.

Abb. 3.5.: Aufbau der Datenbank

Dateiname	Episoden/ Cases	Charakter	Variablen
Rectyp 1	**3777**	**Zeitkonstante Merkmale**	**16**
Rectyp 2	**3777**	**Querschnittdaten**	**59**
Rectyp 3	**10783**	**Allgemeines Kalendarium**	**10**
Rectyp 41	**7576**	**Erwerbskalendarium I**	**36**
Rectyp 42	**7576**	**Erwerbskalendarium II**	**38**
Rectyp 43	**7576**	**Erwerbskalendarium III**	**22**
Rectyp 5	**4986**	**Erwerbslosigkeitskalendarium**	**52**
Rectyp 6	**3777**	**Räumliche Mobilität**	**27**
Rectyp 7	**3777**	**Familienbiographie**	**20**
Rectyp 8	**3777**	**Schulische Ausbildung**	**41**
Rectyp 9	**4148**	**Berufliche Weiterbildung**	**17**
Rectyp 10	**4489**	**Vorgelagerte Zeitepisoden**	**13**
Rectyp 11	**3857**	**Weiterführende Bildungsphasen**	**12**

3.5 Software zur Datenverwaltung und Analyse

Auch wenn die Aufbereitung und Auswertung von Verlaufsdaten (insbesondere mit zeitveränderlichen Kovariaten) mit SPSS längst kein Problem mehr darstellen (Brüderl / Ludwig-Mayerhofer 1994), werden die anstehenden empirischen Analyseverfahren mit dem für Verfahren der Ereignisanalyse flexibleren und umfangreicheren TDA Programm von Götz Rohwer (1994b) durchgeführt. Hierfür mußten die SPSS-Systemdateien nach Abschluß der Eingabearbeiten und Datenkorrekturen in ASCII-Dateien umgewandelt werden. Das gesamte Data Entry Datenbanksystem wurde in ein ZOO-Archiv überführt. ZOO (vgl. hierzu: Rohwer 1992a, 1992b und 1994a) unterstützt wie kein anderes Programm die Organisation von (Zeit-) Daten und den Einsatz zur Analyse von Zeitdaten mittels des Statistikprogramms TDA (vgl. Blossfeld/Rohwer 1995; Rohwer 1994b). Es bietet vielfältige Optionen der Komprimierung, Dekomprimierung und Modifikation und ermöglicht auf relativ unkompliziertem Wege Datenretrivals aus beliebig vielen Dateien in kurzer Zeit.

4 Statistische Evaluation der Stichprobe der Berufsverlaufsstudie Ostdeutschland

Sozialwissenschaftlichen Aussagen, die mit den Methoden der quantitativen empirischen Sozialforschung gewonnen werden, kommt nicht allein deshalb Geltung zu, weil sie auf diesen Methoden basieren. Auch die jeweilige Methode selbst, vor allem aber deren forschungspraktische Anwendung muß bestimmten Gütekriterien gerecht werden. So muß erstens – auf der Ebene der Messung von Merkmalsausprägungen – die Operationalisierung der aus theoretischen Überlegungen abgeleiteten Hypothesen den Kriterien der Validität und der Reliabilität genügen. Und zweitens muß – auf der Ebene der Generierung der Datenbasis – die Auswahl der Untersuchungseinheiten derart gestaltet werden, daß Aussagen über die definierte Gesamtpopulation, d.h. über die Grundgesamtheit, geleistet werden können. Traditionell wird dieses Gütekriterium mit dem Begriff der „Repräsentativität" bezeichnet. Repräsentativität definiert Helmut Kromrey (1998: 259) als

> „Kongruenz zwischen theoretisch definierter Gesamtheit und tatsächlich durch die Stichprobe repräsentierter Gesamtheit; oder: Die Stichprobe ist ein ‚verkleinertes Abbild' einer angebbaren Grundgesamtheit."

Über dieses Konzept der Repräsentativität als Gütekriterium eines Auswahlverfahrens herrscht aber kein Konsens. In den folgenden Ausführungen wird neben diesem Konzept auch ein klassisches Testverfahren zur Überprüfung der Repräsentativität diskutiert und in Anknüpfung an diese Diskussion der Versuch einer Evaluation der Stichprobenziehung der „Berufsverlaufsstudie Ostdeutschland" unternommen.

4.1 Das Konzept der „Repräsentativität"

Damit generalisierbare Aussagen über die soziale Wirklichkeit getroffen werden können, wird von sozialwissenschaftlichen Befragungen erwartet, daß sie, wenn sie nicht die gesamte interessierende Population einschließen, von dieser zumindest eine „repräsentative" Abbildung liefern. So verstehen Peter H. Hartmann und Bernhard Schimpl-Neimanns (1992: 329) unter „Repräsentativität" – im Kern analog zu der Konzeption Kromreys – „die Übereinstimmung der multivariaten Randverteilungen der Merkmale der Stichprobe mit den Verteilungen der Grundgesamtheit". Eine derartige Auffassung ist jedoch nicht unumstritten, da sie, so Ulrich Rendtel und Ulrich Pötter (1993: 351), die Vorstellung von der Stichprobe als ein „Miniaturbild" der Grundgesamtheit impliziere. Diese Vorstellung führe aber insofern zu einem infinitiven Regress, als man auch aus einer Stichprobe eine weitere Stichprobe ziehen können müßte, die – unter der Bedingung, daß sie in allen Merkmalen mit der Grundgesamtheit identisch ist – wiederum ein Miniaturbild der übergeordneten Stichprobe darstellt. Dieser Regress wäre solange weiterführbar, bis man zu einer Stichprobe mit einem „repräsentativen" Menschen kommt. Alle Angaben von Mindestgrößen, mit denen dieser Regress zu stoppen wäre, bleiben dabei willkürlich, wodurch die Miniaturbildkonzeption von Stichproben nicht aufrecht zu erhalten ist.

In einigen neueren Standardwerken zur empirischen Sozialforschung[10] wird schließlich vom Begriff der „Repräsentativität“ Abstand genommen und nur noch von „Zufallsstichprobe“ oder „Wahrscheinlichkeitsauswahl“ gesprochen. Eine realisierte Zufallsstichprobe ist aber weder per se repräsentativ, wie behauptet wird (vgl. Rohwer 1994: 281, Anm. 18), noch sollte sie als Substitution für das Gütekriterium der Repräsentativität betrachtet werden[11]. Vielmehr kann ihre Ziehung unter Verwendung eines Zufallsgenerators theoretisch *jede denkbare* Stichprobe hervorbringen, also auch eine Stichprobe, die – wenn auch mit einer geringen Wahrscheinlichkeit – von der Grundgesamtheit extrem abweichende Merkmalsverteilungen enthält. Würden nämlich theoretisch unendlich viele Stichproben gezogen, ließe sich die Verteilung der Mittelwerte einer Variablen (z.B. Einkommen) als Standardnormalverteilung abbilden und zumindest je 2,5% dieser Mittelwerte würden sich ober- und unterhalb des Wertebereichs von –1,96 bis +1,96 Standardabweichungen vom Mittelwert aller Mittelwerte befinden (vgl. Graff 1992: 130). Zu beachten ist jedoch, daß in sozialwissenschaftlichen Untersuchungen Zufallsstichproben in der Regel nur einmalig gezogen werden und man dabei genau genommen nicht von der Durchführung eines wirklichen Zufallsexperiments ausgehen kann (Rohwer 1994: 278). Werden Zufallsstichproben verwendet, ist die Auswahl der Untersuchungseinheiten von subjektiven Präferenzen der Forschenden bereinigt, was aber andererseits damit erkauft wird, sich auf die Repräsentativität der zufällig ausgewählten Stichprobe verlassen zu müssen (Rohwer 1994: 278f). Schnell/Hill/Esser (1993: 316) halten es zumindest für möglich, mit Hilfe von „Repräsentanznachweisen“ grobe Kontrollen des Ziehungsprozesses durchzuführen. Derartige Nachweise basieren auf Vergleichen der Häufigkeitsverteilungen von Merkmalen in Stichprobe und Grundgesamtheit und setzen damit die Bekanntheit bestimmter Parameter der Grundgesamtheit voraus (vgl. Rendtel 1995: 198, Bedau/Vortmann 1991). Die Grenzen dieser Vergleiche liegen darin, daß nicht bewiesen werden kann, daß im Falle der Gleichverteilung *bestimmter* Merkmale in Stichprobe und Grundgesamtheit auch eine Gleichverteilung hinsichtlich *anderer* Merkmale, über die keine Informationen vorliegen, gegeben ist (Schnell/Hill/Esser 1989: 316, Prein/Kluge/Kelle 1994: 26). Gegen die Annahme, daß bei einer Übereinstimmung der Verteilungen „wichtiger“ Merkmale von Stichprobe und Grundgesamtheit auch alle übrigen (zumeist unbeobachteten) Merkmale übereinstimmen, wenden Rendtel und Pötter (1992: 10) ein:

> „Träfe diese Hypothese zu, so ließe sich die Verteilung der interessierenden sozialen Merkmale aus der bekannten Verteilung der Basismerkmale bestimmen. Wozu dann noch sozialwissenschaftliche Umfragen?“

Es lassen sich darüber hinaus bei komplexeren Datensätzen, die eine Vielzahl von Variablen beinhalten, immer Merkmalskombinationen finden, deren Verteilungen von der Grundgesamtheit abweichen. Werden beispielsweise Repräsentativitätstests hinsichtlich der Merkmale „Geschlecht“, „Alter (12 Kat.)“, „Erwerbsstatus (5 Kat.)“, „Bildung (5 Kat.)“ und „Haushaltsgröße (5 Kat.)“ durchgeführt, erhält man eine Tabelle mit 2*12*5*5*5=3000 Zellen. Auch eine relativ große Stichprobe von n=3000 kann dann praktisch nicht mehr mit den Verteilungen der Grundgesamtheit übereinstimmen. Hinsichtlich komple-

[10] Diekmann 1998: 368, vgl. Rohwer 1994: 281.
[11] Schnell/Hill/Esser (1993: 314) zufolge können die Begriffe „Repräsentativität“ und „Zufallsauswahl“ als synonym betrachtet werden.

xerer Merkmalskombinationen werden folglich auch größere Stichproben kaum repräsentativ sein. Überdies ist auch bei einer perfekt realisierten Zufallsauswahl, die jede Art von systematisch merkmalsbezogenen Selektionsprozessen ausschließt, allein aufgrund des Zufallsfehlers eine Übereinstimmung beider Verteilungen sehr unwahrscheinlich (Rendtel/Pötter 1993: 352).

Die Berufsverlaufsstudie Ostdeutschland wurde als Zufallsstichprobe aus einer spezifisch definierten Grundgesamtheit gezogen, nämlich aus allen Absolventen einer beruflichen Lehre oder eines Hochschulstudiums in Rostock und Leipzig der Jahre 1985, 1990 und 1995. Die Ziehung der Stichprobe stellt somit keine auf die Gesamtbevölkerung abzielende Zufallsauswahl dar, sondern ein quasi- experimentelles Design. Es ist anhand von Vergleichen der Merkmalsverteilungen der *spezifisch* definierten Grundgesamtheit mit den entsprechen Verteilungen der Stichprobe ein Repräsentanznachweis für jene Merkmale möglich, deren relative Häufigkeitsverteilungen in der Grundgesamtheit bekannt sind[12]. Man hat damit zumindest die Möglichkeit einer Kontrolle, ob die Ziehung der Stichprobe signifikant mit systematischen Abweichungen hinsichtlich der geprüften Merkmale einherging. Derartige Datenvergleiche liefern daher zumindest einen Anhaltspunkt, ob gröbste Verletzungen der Zufallsauswahl vorliegen. Man sollte deshalb trotz der genannten Probleme auf deren Durchführung nicht verzichten.

4.2 Systematische Ausfälle durch Non-response und Panelmortalität

Obwohl die Auswahl der befragten Personen aus den vorliegenden Adressenbeständen mit Hilfe eines Zufallsgenerators vollzogen wurde und eine Zufallsauswahl damit gegeben ist, können Verzerrungen entstanden sein, wenn die Ausfälle durch Teilnahmeverweigerung, „Item-Nonresponse“[13] oder Nicht-Erreichbarkeit nicht zufallsverteilt, sondern mit Merkmalen wie Geschlecht, Berufstätigkeit und Alter korreliert sind (Prein/Kluge/Kelle 1994: 13). Bei postalischen Befragungen treten non- response- Ausfälle besonders häufig auf (Kromrey 1998: 378) und unter Umständen sind diese Ausfälle in Abhängigkeit vom Bildungsgrad oder des Ausbildungsberufs der Befragten zu erwarten.

Die Übertragbarkeit der Stichprobenergebnisse auf die Grundgesamtheit ist in der *Längsschnittperspektive* darüberhinaus mit dem Problem der Panelmortalität konfrontiert. Wie auch die Antwortverwei-

[12] Die Informationen über die Grundgesamtheit für die Hochschulabsolventen in Leipzig 1995 liegen jedoch in einer Form vor, die eine Einschätzung des Nettosamples hinsichtlich systematischer Ausfälle nicht unmittelbar zuläßt. So sind in dieser Grundgesamtheit 227 Personen als „Lehrer“ aufgeführt, im Nettosample wurden diese jedoch aufgrund ihrer Fachrichtung codiert. Überdies trat das Problem auf, daß im Nettosample 6 Ingenieure existieren (Tab. 4.7), obwohl laut Angaben des Statistischen Bundesamtes (zumindest nach formaler Definition) in Leipzig keine Ingenieure ausgebildet wurden. Zu bedenken ist hierbei, daß wegen der Datenschutzbestimmungen auf der Ebene der Individualdaten die Fachrichtungen der Abschlüsse auf den Selbsteinschätzungen der Befragten basieren mußten. Es war aber möglich, die Berufsgruppen mittels einer Verknüpfung von Fachrichtung und StaBu- Berufsklassifikation des Erstberufs nachzukodieren und die Personen des Samples somit relativ eindeutig den Berufsgruppen zuzuweisen. Die erste Anwendungsvoraussetzung, die nach Bortz (1989: 197) für die Durchführung des χ^2 – Anpassungstests erfüllt sein muß, daß „jede untersuchte Einheit eindeutig einer Kategorie zugeordnet werden kann“, ist somit erfüllt. Durch die durchgeführten Umkodierungen sind die Abweichungen zwischen den Tabellen 4.1 und 4.2 und den Tabellen 4.4 bis 4.8 zu erklären.

[13] Dabei handelt es sich um den Fall, daß bei bestimmten Variablen Angaben fehlen (Schnell/Hill/Esser 1993: 316). Betroffen sind davon bspw. insbesondere Auswertungen von Einkommensangaben, wenn die Antwortverweigerung systematisch mit der Höhe des Einkommen korreliert ist.

gerungen in der ersten Erhebungswelle kann die Panelmortalität mit bestimmten Merkmalen der Befragten korreliert sein (Diekmann 1998: 271, Rendtel 1995)[14].

4.3 Was ist die Grundgesamtheit?

Für die Evaluation der Repräsentativität der SFB-3 Lebensverlaufsstudie zog Hans-Peter Blossfeld (1989: 101) Merkmale wie Alter, Geschlecht, Stellung im Beruf heran, die ebenfalls in den Mikrozensen 1981-1982 sowie in der Mikrozensus- Zusatzerhebung von 1971 erhoben wurden. Auch Hartmann/Schimpl-Neimanns (1992) definierten den Mikrozensus als Grundgesamtheit, da bei dessen Ziehung das non- response Problem durch die Antwortverpflichtung der Befragten minimiert sei und die Fallzahlen sehr hoch sind. Rendtel/Pötter (1992: 10) gehen demgegenüber davon aus, daß zum einen auch die Mikrozensusergebnisse einer stichprobenbedingten Zufälligkeit unterliegen. Zum anderen weisen sie auf Abweichungen der Haushaltsgrößenverteilungen zwischen dem Mikrozensus 1987 und der nur drei Monate später durchgeführten Volkszählung hin. So war im Vergleich zur Volkszählung im Mikrozensus die Zahl der Einpersonenhaushalte um 600.000 erhöht. Erklärt wird diese Abweichung mit konträren Anreizen für die Interviewer beider Erhebungen. Beim Mikrozensus werden die Interviewer nach der Anzahl der befragten Haushalte bezahlt und sind darum geneigt, eine vorgefundene Haushaltskonstellation möglichst so zu interpretieren, als bestünde sie aus mehreren kleinen separaten Haushalten. Interviewer der Volkszählung hatten diese Anreize hingegen nicht und waren eher bestrebt, Mehrarbeit durch Anlegen eines neuen Haushaltsmantelbogens zu vermeiden und damit Konstellationen kleinerer Haushalte zu wenigen großen Haushalten zusammenzufassen (ebd.: 11). Wird nun der Mikrozensus als Vergleichsdatensatz für einen Repräsentanznachweis herangezogen, muß beachtet werden, daß auch dieser hinsichtlich der Verteilungen der Haushaltsgröße und eventuell korrelierender Variablen nicht zwangsläufig „die" Grundgesamtheit widerspiegelt.

Noch gravierender als bei der Lebensverlaufsstudie des SFB-3 ist bei der Berufsverlaufsstudie Ostdeutschland die Tatsache, daß sie einer noch weitaus spezielleren Teilpopulation entstammt und der Mikrozensus nicht für einen Abgleich der relativen Häufigkeiten herangezogen werden kann. Dafür stehen aber die einzelnen Grundgesamtheiten aufgrund der engen Definition (drei Absolventenkohorten, zwei Ausbildungsorte, zwei Bildungsabschlüsse) eindeutig fest. Und weil die Adressenbestände der Grundgesamtheit aufgeschlüsselt nach Fachrichtung und Berufsgruppen nahezu vollständig vorlagen, ist zumindest eine Kontrolle der Verteilung dieser Merkmale möglich. Insbesondere durch sie sind nämlich Selektionseffekte aufgrund von non- response zu erwarten. Wegen des in vielen Umfragen auftretenden „Mittelschichtsbias" (Hartmann/Schimpl-Neimanns 1992: 324ff, Rendtel 1995: 194, Diekmann 1998: 361) wird vermutet, daß die Teilnahmebereitschaft an Befragungen insbesondere bei Personen aus niedrigen Bildungs- und Einkommensschichten geringer ist, als jene von Personen der „Mittelschicht". Das dieses Phänomen auch in dieser Studie aufgetreten ist, wurde bereits ausgeführt (Abschnitt. 2.6). Sollte

[14] Die Effekte der Panelmortalität sind in unserem Fall evaluierbar, indem auch für die späteren Erhebungswellen Vergleiche von Verteilungen der Stichprobe und der Grundgesamtheit durchgeführt werden und nach systematischen Ausfällen gesucht wird (vgl. Rendtel 1995: 196f).

neben der größeren sozialen Distanz der Lehrabsolventen zu den von den Akademikern geleiteten sozialwissenschaftlichen Umfragen auch der berufliche Erfolg die Teilnahmebereitschaft beeinflussen, müßte sich dies auch in den folgenden Analysen widerspiegeln. Demzufolge ist eine Überrepräsentation von Berufsgruppen und Fachrichtungen zu erwarten, deren Humankapitalressourcen besser an die nach der Wende gewandelte Nachfrage angepaßt war, bzw. eine Unterrepräsentation von Gruppen, die von einer Entwertung ihrer Qualifikation betroffen waren. Der Wandel der Nachfrage bestand insbesondere in der „nachholenden Tertiarisierung" der Beschäftigungsstruktur, wodurch unter der oben anführten Annahme Fachrichtungen und Berufsgruppen sich in ihrem Antwortverhalten je nach Nähe zum tertiären Sektor unterscheiden.

4.4 Der χ^2 - Anpassungstest

Ist die Verteilung einer Variablen in der Grundgesamtheit bekannt, können anhand der in der Stichprobe vorliegenden Fallzahlen die aufgrund der Populationsverteilung erwarteten Häufigkeiten errechnet werden, indem der prozentuale Anteil einer Kategorie der Grundgesamtheit durch 100 dividiert und mit der Fallzahl der Stichprobe multipliziert wird[15]. Weil aber jede Stichprobenziehung mit Zufallsschwankungen einhergeht, ist eine Abweichung der beobachteten Verteilung von der erwarteten Verteilung zu erwarten. Mit Hilfe des χ^2–Anpassungstests kann festgestellt werden, ob diese Abweichung innerhalb des durch die Zufallsschwankungen bedingten Rahmens liegt oder ob sie durch systematische Ausfälle begründet ist[16]. Über die Aufsummierung der (quadrierten) und durch den Erwartungswert dividierten Differenzen von beobachteter und erwarteter Verteilung erhält man eine χ^2–verteilte Testgröße:

$$c^2 = \sum_{j=1}^{k} \frac{(f_{b(j)} - f_{e(j)})^2}{f_{e(j)}}$$

Unter Rückgriff auf die χ^2– Verteilungstabelle kann nun ermittelt werden, ob der jeweilige χ^2– Wert bei gegebenen Freiheitsgraden (Bortz 1989: 198) signifikant ist.

Eine wichtige Frage stellt dabei die Wahl des Schwellenwertes dar, anhand dessen das Ausmaß einer Abweichung des samples von der Verteilung in der Grundgesamtheit festgestellt wird (Rendtel 1995: 201). So beurteilt Blossfeld (1989: 110) die Vergleichbarkeit der SFB-3-Lebensverlaufsstudie mit den Daten des Mikrozensus in bezug auf die prozentualen Verteilungen der Befragten über die Bundesländer als „gute Anpassung". Seine Begründung besteht darin, daß nur eine Kategorie einen χ^2– Wert über der kritischen Schwelle der 1% Irrtumswahrscheinlichkeit aufweist (23,2 bei df=10).

Bei diesem Vorgehen ist aber grundsätzlich zu bedenken, daß eine derartig konservative Schwelle nur hinsichtlich einer Hypothese über den *Zusammenhang* zwischen zwei Merkmalen (z.B. Geschlecht und Parteipräferenz) als sinnvoll erachtet werden kann, *nicht* aber bei einen „Repräsentativitätstest". Konservativ ist ein Schwellenwert dann, wenn strenge Kriterien erfüllt sein müssen, damit die Nullhypothese verworfen werden kann (vgl. Diekmann 1998: 594). Die Wahrscheinlichkeit, daß ein in der Stichprobe festgestellter *Zusammenhang* zwischen zwei Merkmalen rein zufallsbedingt ist, wird dadurch minimiert.

---

[15] formal: $f_e = n * p_j$, wobei f_e die erwartete Häufigkeit des Merkmals j im Sample, p die relative Häufigkeit des Merkmals in der Grundgesamtheit und n die Samplegröße darstellt (Bortz 1989: 189).

[16] kritisch gegenüber diesem Vorgehen ist Rendtel (1995: 199ff).

Ist bei einer Zusammenhangsanalyse eine konservative Schwelle gesetzt, wird zugleich eine hohe Wahrscheinlichkeit des - Fehlers [17] hingenommen. Es steigt also die Gefahr, einen in der Grundgesamtheit durchaus existierenden Zusammenhang zwischen zwei Merkmalen zu „übersehen", weil das Prüfkriterium für die Akzeptanz des Zusammenhangs zu streng formuliert wurde. Anders ausgedrückt: bei einer konservativen Schwellensetzung müssen sich die Befunde aus der Stichprobe schon sehr eindeutig abzeichnen, damit von einem Zusammenhang zwischen den Variablen gesprochen werden kann. Die Hürden für eine Annahme neuer Erkenntnisse sind also höher gesetzt.

Wird dagegen bei einem *Repräsentativitätstest* eine Sample- Verteilung mit einer bekannten Verteilung der Grundgesamtheit verglichen, um systematische Abweichungen einer Stichprobenziehung von der reinen Zufallsauswahl aufzudecken, ist es wenig ratsam, konservative Schwellen zur Identifikation einer Abweichung der Verteilungen zu setzen. Die bei der Prüfung der „Repräsentativität" eingenommene Perspektive ist nämlich die eines „goodness of fit tests", mit dem bspw. überprüft wird, ob die empirisch vorgefundene Verteilung signifikant von einer theoretisch erwarteten Verteilung abweicht (Bortz 1989: 197). In dieser Perspektive ist es angebracht, die Prüfkriterien für den erwünschten Befund einer *Übereinstimmung* – nicht einer Differenz – beider Verteilungen möglichst streng zu formulieren. Denn es wird ja eine Aufrechterhaltung der Nullhypothese angestrebt, nicht deren Verwerfung[18]. Mit der Minimierung des - Fehlers bei einem Repräsentativitätstest, durch die eine steigende statistische Absicherung einer Aussage über die mögliche Differenz beider Verteilungen erreicht wird, sinkt zugleich die Wahrscheinlichkeit, daß diese Differenzen festgestellt werden. Denn je konservativer die Schwellensetzung ist, desto eindeutiger müssen sich die Abweichungen abzeichnen, damit sie als signifikant ausgewiesen werden können. Mit der Minimierung des - Fehlers ist der Test der Anpassung einer Stichprobe an die Merkmalsverteilungen der Grundgesamtheit also *weniger* strengen Prüfkriterien ausgesetzt!

Bortz (ebd. 198) betont darum, daß „nicht die Wahrscheinlichkeit des - Fehlers, sondern die Wahrscheinlichkeit des - Fehlers möglichst klein sein sollte". Indirekt kann der - Fehler klein gehalten werden, indem man den maximal erlaubten - Fehler *vergrößert*. Auf diese Weise werden auch schon geringere Abweichungen zwischen den Verteilungen als Widerlegung der Nullhypothese gewertet.

Verkompliziert wird die Frage des Schwellenwertes überdies durch den Sachverhalt des *multiplen Testens*. Wird eine Vielzahl von Signifikanztests voneinander unabhängiger Nullhypothesen durchgeführt, steigt auch der - Fehler, also die Wahrscheinlichkeit, fälschlicherweise signifikante Differenzen zu konstatieren. So tritt bei den 36 Tests, die Hartmann/Schimpl-Neimanns (1992: 329, *Tab.* 6) bei einer Irrtumswahrscheinlichkeit von 1% als Vergleiche von Allbus- und Mikrozensuserhebungen durchführten, mit einer Wahrscheinlichkeit von 30% mindestens eine nur durch den Zufallsfehler bedingte Abwei-

---

[17] - Fehler bedeutet, daß aufgrund der Befunde aus der Stichprobe die H_0 (die Nullhypothese: „Es besteht *kein* Zusammenhang zwischen den Merkmalen") fälschlicherweise aufrecht erhalten wird, obwohl in der Grundgesamtheit die H_1 („es besteht ein Zusammenhang zwischen den Merkmalen") zutrifft (Bortz 1989: 145, Rendtel 1995: 201, Graff 1992: 136).

[18] Hartmann/Schimpl-Neimanns (1992: 316, Anm. 3) formulieren dieses Erkenntnisinteresse unverblümt: „Der Vergleich der Umfragen mit den amtlichen Daten dient meist dem Zweck, die Nutzung des Umfragematerials für spezielle inhaltliche Untersuchungen zu legitimieren".

chung[19] auf (Rendtel/Pötter 1993: 353). Jürgen Kritz (1973: 116) empfiehlt darum auch eine äußerst sparsame Verwendung von Signifikanztests.

4.5 Repräsentanztest der Berufsverlaufsstudie Ostdeutschland

Die vorangegangenen Ausführungen dienten vor allem als Hinweis auf die Problematik des klassischen Vorgehens bei Repräsentativitätsprüfungen nach dem χ^2 – Anpassungstest. Im folgenden wird dieser Test mangels Alternative (Hartmann/Schimpl-Neimanns 1993: 364) dennoch durchgeführt, um einen Eindruck zu gewinnen, wie die non- response bedingten Antwortausfälle nach dem klassischen Testverfahren einzuschätzen sind. Schließlich ist Prein/Kluge/Kelle (1994: 27) zuzustimmen, wenn sie folgern:

> „Angesichts des Fehlens methodologischer Alternativen ist also beim Einsatz von Datenvergleichen und Anpassungstests zwar extreme Vorsicht geboten, wenn hieraus die generelle Unverzerrtheit von Datensätzen abgeleitet werden soll; es erscheint jedoch uneinsichtig, warum ein Verfahren, das in bestimmten Situationen in der Lage ist, Verzerrungen aufzudecken, in der Forschungspraxis keine Anwendung finden sollte."

Die χ^2 –Anpassungstests[20] weisen auf signifikante Abweichungen zwischen den Verteilungen der Grundgesamtheit und der Stichprobe hin, wenn der jeweilige χ^2 - Wert einen dem jeweils gewählten Signifikanzniveau entsprechenden Schwellenwert überschreitet. Die durchgeführten Tests lieferten dabei folgende Ergebnisse:

---

[19] Sind die Tests voneinander unabhängig, steigt der - Fehler nicht proportional mit der Zahl der Tests, so ndern wird durch einen speziellen, aber einfachen Ausdruck adjustiert (Rendtel/Pötter 1992: 9, Bortz 1989: 321): '=1-(1-) ^{m}, wobei ' den für das multiple Testen adjustierten - Fehler darstellt, das bei den ei nzelnen Tests akzeptierte Irrtumswahrscheinlichkeit und m die Anzahl der durchgeführten Signifikanztests.

[20] Zu beachten sind bei der Durchführung des χ^2 –Anpassungstests die von Bortz (1989: 197) genannten Anwendungsvoraussetzungen. So müssen die Erwartungshäufigkeiten in jeder Kategorie größer als 5 sein (vgl. auch Prein/Kluge/Kelle 1994: 49). Weil das bei nicht bei allen Populationen der Fall war, wurden einige Zusammenfassungen vorgenommen (vgl. Bortz 1989: 197, vgl. hierzu auch die Tab. 4.1-4.4 mit den Tab. 4.5-4.6 im Anhang):

Bei den ***Hochschulabsolventen in Leipzig*** wurden

- bei der Kohorte 1985 Ingenieure (N=21 GG, n=2 Sample) zu den Naturwissenschaftlern (N=220 GG, n=42 Sample) und „Kunst, Kunstwissenschaften" (N=36 GG, n=4 Sample) zu „Sprach- und Kulturwissenschaften" (N=529 GG, n=75 Sample) gezählt.
- bei der Kohorte 1990 Ingenieure (N= 20 GG, n=3 Sample) zu den Naturwissenschaftlern (N=20 GG, n=3 Sample) und die Absolventen von Kunst und Kunstwissenschaft (N=18 GG, 4 Sample) der Kategorie „Sprach- und Kulturwissenschaft" (N=533, n=107) zugeordnet, und
- bei der Kohorte 1995 Absolventen von Kunst und Kunstwissenschaft (N=11 GG, n=2 Sample) und Absolventen der Sportwissenschaften (N=88 GG, 2 Sample) in die Kategorie der Lehrer eingefügt.

Bei ***den Lehrabsolventen in Leipzig*** wurden

- bei der Kohorte 1985 die Pflanzenbauer/ Tierzüchter/ Fischer (N=53 GG, n=0 Sample) den technischen Berufen (N=218 GG, n=11 Sample) zugewiesen.
- Gleiches gilt für die Kohorte 1990 (Pflanzenbauer/ Tierzüchter/ Fischerei N=80 GG, n=0 Sample, technische Berufe N=187 GG, n=20 Sample).
- Bei der Kohorte 1995 wurden die technischen Berufe (N=22 GG, n=3 Sample) den Fertigungsberufen (N=1582 GG, n= 131 Sample) zugewiesen.

Bei den ***Lehrabsolventen in Rostock*** wurde

- bei der Kohorte 1985 die Kategorie Pflanzenbauer/ Tierzüchter/ Fischerei (N=81 GG, n=0 Sample) unter die technischen Berufen (N=89 GG, n=4 Sample) subsummiert.
- Gleiches geschah bei der Kohorte 1990 (technische Berufe N=55 GG, n=4 Sample, Pflanzenbauer/ Tierzüchter/ Fischerei N= 100 GG, n=0 Sample).

4.5.1 Lehrabsolventen

Bei den Lehrabsolventen trat bei 5 von 6 Tests keine signifikanten Abweichungen auf. Bei den Leipziger Lehrabsolventen des Jahres 1985 zeigte sich dagegen eine Überrepräsentation der Dienstleistungsberufe und eine Unterrepräsentation der Fertigungsberufe (Tab. 4.1. u. 4.2). Bei nur einer von sechs signifikanten Abweichungen kann aber noch nicht von einer generellen systematischen Überrepräsentation der gegenüber den Fertigungsberufen nachfrageadäquater ausgebildeten Dienstleistungsberufe ausgegangen werden. In Abschnitt 4.3 wurde dies ja als mögliche Quelle für systematische Abweichungen angeführt.

4.5.2 Hochschulabsolventen

Tab. 4.3 und *4.4* stellen die χ^2 –Anpassungstests für die Hochschulabsolventen in Leipzig dar. Im Gegensatz zu den beiden anderen Kohorten weist die Kohorte 1985 bei einer 5% Irrtumswahrscheinlichkeit keine signifikanten Abweichungen auf. Das Fehlen von signifikanten Abweichungen indiziert in diesem Zusammenhang den erwünschten Befund der „Repräsentativität". Liegen signifikante Abweichungen vor, dann weichen die Verteilungen der Stichprobe von jenen der Grundgesamtheit ab, die „Repräsentativität" ist damit in Frage gestellt. Auffällig ist der überhöhte χ^2 – Wert der Kohorte 1995, der vor allem aus einer Überrepräsentation der Fachrichtungen „Rechts-, Wirtschafts- und Sozialwissenschaften" und einer Unterrepräsentation der Fachrichtungen „Mathematik/ Naturwissenschaften" und „Agrar-, Forst- und Ernährungswissenschaften" resultiert. Es ist an dieser Stelle zu betonen, daß die auf einer Selbsteinstufung der Befragten basierende Angabe im Fragebogen durchaus mit den formalen Definitionen der Fachrichtungen an den Universitäten abweichen kann. Das könnte bedeuten, daß die Abweichungen zwischen den Verteilungen nicht allein aus dem Responseverhalten resultieren, sondern auch mit der unterschiedlichen Definition der Abschlüsse durch die Ausbildungsanstalten einerseits und die Befragten andererseits zusammenhängen.

Nicht signifikant sind die Abweichungen der beiden Rostocker Kohorten 1990 und 1995 (Tab. 4.4). Dagegen sind bei der Kohorte 1985 Abweichungen aufgrund der Unterrepräsentation der Kategorien „Sport" und „Rechts-, Wirtschafts- und Sozialwissenschaften" sogar auf dem 1% - Niveau signifikant.

Letztere Kategorie war bei der Leipziger Kohorte 1995 aber nicht unter- sondern überrepräsentiert. Von einer generellen fachrichtungsbedingten Systematik kann bei den festgestellten Abweichungen im Antwortverhalten auch bei den Akademikern folglich nicht ausgegangen werden.

Insgesamt läßt sich sagen, daß beim Vergleich der Anpassung der Verteilungen zwischen Stichprobe und Grundgesamtheit in Einzelfällen durchaus Abweichungen von der statistischen Unabhängigkeit aufgetreten sind.

Betrachtet man aber alle Stichprobenziehungen der Berufsverlaufsstudie Ostdeutschland zusammen, beschränkt sich die Zahl der bei einer 5%- Irrtumswahrscheinlichkeit von den Verteilungen der Grundgesamtheit abweichenden Ziehungen auf 4 von 12, bei einer 1% Irrtumswahrscheinlichkeit sogar nur auf 3 von 12. Die durch non- response bedingten Verzerrungen werden sich damit wohl in Grenzen halten[21].

[21] Die Wahrscheinlichkeit, daß eine der vorgefundenen Abweichungen allein durch den Zufallsfehler bedingt ist, beträgt überdies ‘=1-(1-) ^{m} = 1-(1-0,05)^{12} = 46%.

Die einzige wirklich auffällige Abweichung zwischen den beobachteten und den bei statistischer Unabhängigkeit erwarteten Häufigkeiten trat bei den Leipziger Hochschulabsolventen von 1995 auf. Dies ist aber in sozialwissenschaftlichen Untersuchungen nicht unüblich. Ähnlich hohe ^{2}- Werte lassen sich bspw. auch in den Tests der SFB-3 Lebensverlaufsstudie finden (Blossfeld 1989: 115), bei der immerhin die Möglichkeit gegeben war, die non- response Quote durch den Einsatz von Interviewern zu minimieren. Ob derartige Abweichungen zu verzerrten Modellschätzungen führen können, kann ad hoc nicht festgestellt werden. Rendtel (1995: 202) betont aber in diesem Zusammenhang:

> „Der negative Ausgang einer Repräsentativitäts- Prüfung über den Vergleich mit vorgegebenen Populationswerten bedeutet (..) nicht, daß Populationsschätzungen mit der gezogenen Stichprobe unmöglich sind."

Ein positiver Befund einer Untersuchung auf Verteilungsabweichungen offenbare vielmehr, daß der Prozeß der Datenerhebung anders verlaufen ist, als erwartet wurde. Die Korrektur der Abweichungen müsse über zu ändernde Schätzverfahren erfolgen.

Schließlich ist anzumerken, daß die Erhebung der Berufsverlaufsstudie Ostdeutschland als quasi- experimentelles Design angelegt war (Windzio 2000). Angestrebt wurde also keine repräsentative Querschnittsstichprobe, sondern eine auf a priori formulierten theoretischen Annahmen basierende Unterscheidung in Experimental- und Kontrollgruppe. Dieses Vorgehen ist angemessen, wenn das Ziel der Untersuchung in einer Prüfung von Hypothesen besteht, während repräsentative Querschnittsstichproben eine Voraussetzung für deskriptive Analysen darstellen (Diekmann 1998: 169). Durch non- response bedingte Ausfälle verlieren in einem quasi- experimentellen Design etwas an Dramatik, wenn die einzelnen, für die Definition als Versuchs- und Kontrollgruppe relevanten Subpopulationen in der Stichprobe hinreichend besetzt sind.

Hartmann/Schimpl-Neimanns (1992: 337) betonen darüberhinaus – nachdem sie teilweise deutliche Abweichungen der Allbus Daten von Daten des Mikrozensus hinsichtlich der Variablen Haushaltsgröße, Bildung und Erwerbstätigkeit feststellten –, daß diese Abweichungen bei deskriptiven Analysen natürlich verzerrte Randverteilungen zur Folge haben, wenn die betroffenen Merkmale nicht kontrolliert sind. Im „Fall der bedingten Regression" sei dagegen nicht mit systematischen Verzerrungen zu rechnen.

Untersucht wurden die non- response Effekte in den vorliegenden Analysen aber nur in bezug auf die Startstichprobe des Makro- Panels. Zu fragen bleibt, inwieweit auch die Panelmortalität mit systematischen Unterschieden zwischen den einzelnen Subpopulationen einher geht. Zunächst ist festzuhalten, daß im Gegensatz zur Startstichprobe Datenvergleiche zwischen den einzelnen Wellen schon allein darum leichter durchzuführen sind, weil nun auch Merkmale herangezogen werden können, die bei der Erhebung erfragt wurden. Allerdings werden natürlich alle Ausfälle der Startstichprobe fortgeführt, so daß sich eine Evaluation der letzten Panelwelle, deren Aussagekraft für die Grundgesamtheit sowohl durch non- response Effekte der Startstichprobe als auch durch die Panelmortalität beeinträchtigt sein kann, eher auf die Verteilung der Grundgesamtheit der Startstichprobe beziehen sollte. Erste deskriptive Befunde zur Panelmortalität der Berufsverlaufsstudie Ostdeutschland in der zweiten Welle liegen in Kap. 6 vor, abschließende Analysen erfolgen nach der abschließenden, dritten Panelwelle im Jahr 2000.

4.6 Schlußfolgerung

Wie anhand der jüngeren Diskussion gezeigt wurde, wirft das Konzept der „Repräsentativität“ bei genauerer Reflexion einige Probleme auf. Eine einfache Substitution dieses Gütekriteriums der Datenerhebung durch die Methode der Zufallsauswahl erscheint nicht empfehlenswert, da Zufallsgeneratoren auch extrem von der Grundgesamtheit abweichende Stichproben ziehen können. Eine Zufallsauswahl ist noch keine Garantie für eine Kongruenz von Stichprobe und Grundgesamtheit in den Merkmalsverteilungen. Darüberhinaus können non- response bedingte Ausfälle systematisch mit bestimmten Merkmalen der Befragten einhergehen. Darin lag der Grund, den χ^2 – Anpassungstest auch zu einer Evaluation der Zufallsstichprobe der Berufsverlaufsstudie Ostdeutschland heranzuziehen. Werden aufgrund des Tests gröbste Abweichungen festgestellt, ist die Güte und die Aussagefähigkeit der Daten in bezug auf die interessierende Gesamtpopulation eingeschränkt. Werden keine signifikanten Abweichungen festgestellt, heißt das dagegen nicht zwangsläufig, daß Stichprobenziehung und non- response Ausfälle in jeglicher Hinsicht unproblematisch sind.

Die 12 durchgeführten Tests liefern zusammen ein recht befriedigendes Bild des Rücklaufprozesses der postalischen Befragung. Die vorgefundenen Abweichungen sind vergleichbar mit denen der SFB-3 Lebensverlaufsstudie, die von deren Primärforschern auch insgesamt als sehr erfolgreich eingestuft wurde (Blossfeld 1989). Bemerkenswert ist aber, daß das Responseverhalten in der Berufsverlaufsstudie Ostdeutschland bei den Lehrabsolventen etwas weniger mit den einzelnen beruflichen Abschlüssen korreliert war, als mit den Fachrichtungen der Hochschulabsolventen, obwohl im allgemeinen erwartet wird, daß „die Kooperationsbereitschaft mit dem Bildungsgrad ansteigt“ (Diekmann 1998: 361).

Wichtig ist bei der Anwendung des ^{2} – Anpassungstests, daß immer die Möglichkeit von - und - Fehlern beachtet werden. So besteht zum einen auch bei einer perfekt realisierten Zufallsauswahl die Möglichkeit, signifikante Abweichungen zwischen den Verteilungen in Stichprobe und Grundgesamtheit vorzufinden (- Fehler). Denn mit einer gewissen Wahrscheinlichkeit liefert auch eine Zufallsauswahl extreme Verteilungen in der Stichprobe, weshalb im Falle des Befundes signifikanter Abweichungen die Wahrscheinlichkeit deren Zufälligkeit geprüft werden sollte. Zum anderen besteht aber auch die Möglichkeit des - Fehlers. Das bedeutet, daß der ^{2} – Test zwar keine signifikante Abweichung anzeigt, dieser Befund aber dadurch bedingt ist, daß ein Signifikanzniveau von 5% zu konservativ und die kritische Schwelle damit zu hoch angesetzt ist, um diese Abweichungen aufzudecken. Ein derartiger Test, wie er auch hier durchgeführt wurde, besagt beispielsweise, daß ein ^{2} – Wert von 7.56 bei 5 Freiheitsgraden auf keine signifikante Abweichung der Verteilungen der Stichprobe von denen der Grundgesamtheit hindeutet. „Signifikanz“ bedeutet aber auf dem 5%-Niveau, daß bei 95 von 100 gezogenen Stichproben ein Unterschied in den Verteilungen festzustellen ist. Würde bei 6 Stichproben von 100 eine Gleichverteilung auftreten, würde man den Unterschied nicht mehr als signifikant betrachten. Angesichts der Tatsache, daß ein Unterschied aber bei den restlichen 94 Stichproben festzustellen ist, sollte man angesichts des hohen - Fehlers nicht voreilig davon ausgehen, daß keine Abweichungen zwischen Stichprobe und Grundgesamtheit existieren. Immerhin beträgt der - Fehler eines ^{2} – Werts von 7.56 bei 5 Freiheitsgraden auf dem 5%-Niveau ganze 46% (!), d.h. die Wahrscheinlichkeit, aufgrund des

Tests eine richtige Aussage zu treffen ist damit unwesentlich größer, als diejenige einer falschen Aussage.

Als Fazit ist darum festzuhalten, daß zwar Methoden zur Evaluation der Güte einer Stichprobe existieren. Sie sind aber sehr begrenzt in ihrer Aussagekapazität. Man kann durch sie immerhin Hinweise erhalten, ob massive Abweichungen zwischen den Verteilungen in Grundgesamtheit und Stichprobe auftreten (wie es der Fall ist bei den Leipziger Hochschulabsolventen der Kohorte 1995). Wie damit in der konkreten Auswertung der Daten zu verfahren ist, hängt einerseits von den angewandten Analyseverfahren ab: während deskriptive Auszählungen problematisch sind, werden kausale Analysen nicht grundsätzlich von diesen Abweichungen tangiert – insbesondere dann nicht, wenn die Erhebung ohnehin als quasi- experimentelles Erhebungsdesign angelegt ist.

Tab. 4.1: χ^2 –Anpassungstests für Lehrabsolventen in Rostock

	Rostock 1985			Rostock 1990			Rostock 1995		
Fächergruppe	GG	Netto-samp.	erwartete Häufigkeit	GG	Netto-samp.	erwartete Häufigkeit	GG	Netto-samp.	erwartete Häufigkeit
1. Pflanzenbauer/ Tierzüchter/ Fischerei	--	--	--				--		--
2. Bergleute/ Mineralgewinner	--	--	--	--	--	--	--	--	--
3. Fertigungsberufe	3100 55.75%	107 53.5%	111.5	2185 48.62%	116 44.91%	114.74	696 42.28%	115 42.27%	115
4. technische berufe	170 3.05%	4 2%	6.1	155 3.44%	4 1.69%	8.1	31 1.88%	6 2.2%	5.11
5. Dienstleistungsberufe	2290 41.18%	89 44.5%	82.36	2154 47.93%	116 49.15%	113.11	919 55.83%	151 55.51%	151.85
Gesamt	5560 100%	200 100%	200	4494 100%	236 100%	236	1646 100%	272 100%	272
χ2-Wert df	1.43 n.s. 2			2.16 n.s. 2			0.15 n.s. 2		
	0.181 0.722 0.535 1.43			0.013 2.075 0.073 2.161			0 0.155 0.0047 0.159		

* signifikant (p<=0.05) ** hoch signifikant (p<=0.01) *** höchst signifikant (p<=0.001)

Quelle: Berufsverlaufsstudie Ostdeutschland, eigene Berechnungen

Tab. 4.2: χ^2 –Anpassungstests für Lehrabsolventen in Leipzig

	Leipzig 1985			Leipzig 1990			Leipzig 1995		
Fächergruppe	GG	Netto-samp.	erwartete Häufigkeit	GG	Netto-samp.	erwartete Häufigkeit	GG	Netto-samp.	erwartete Häufigkeit
1. Pflanzenbauer/ Tierzüchter/ Fischerei							--	--	--
2. Bergleute/ Mineralgewinner	--	--	--	--	--	--			--
3. Fertigungsberufe	3748 67.06%	109 57.98%	126.07	3553 63.2%	159 63.34%	158.63	1604 48.99%	134 43.64%	150.4
4. technische berufe	271 4.84%	11 5.85%	8.42	187 3.32%	4 1.59%	8.33	--	--	--
5. Dienstleistungsberufe	1579 28.25%	68 36.17%	53.11	1917 34.1%	88 35.06%	85.6	1670 51.0%	173 56.35%	156.57
Gesamt	5589 100%	188 100%	188	5621 100%	251 100%	251	3274 100%	307 100%	307
χ2-Wert df		7.27* 2			2.31 n.s. 2			3.512 n.s. 1	
		2.311			0			1.788	
		0.790			2.250			1.724	
		4.174			0.067				
		7.275			2.317			3.512	

* signifikant (p<=0.05) ** hoch signifikant (p<=0.01) *** höchst signifikant (p<=0.001)

Quelle: Berufsverlaufsstudie Ostdeutschland, eigene Berechnungen

Tab. 4.3: χ^2 –Anpassungstests für Hochschulabsolventen in Leipzig

	Leipzig 1985			Leipzig 1990			Leipzig 1995		
Fächergruppe	GG	Netto-samp.	erwartete Häufigkeit	GG	Netto-samp.	erwartete Häufigkeit	GG	Netto-samp.	erwartete Häufigkeit
1 Sprach-/ Kulturwissenschaften	565 32.73	79 32.91	78.55	551 34.05	111 33.63	112.36	186 10.36	57 11.15	52.93
2 Sport	--	--	--	--	--	--	--	--	--
3 Rechts-/ Wirtschafts-/ Sozialwissenschaften	298 17.26	39 16.25	41.42	237 14.84	39 11.82	48.97	287 15.98	173 33. 85	81.65
4 Mathematik, Naturwissenschaften	241 13.96	44 18.33	33.50	219 13.71	67 20.30	45.24	447 24.9	66 12.91	127.23
5 Humanmediziner	397 23	42 17.5	55.2	369 23.1	63 19.09	76.23	388 21.61	91 17.8	110.42
6 Veterinärmedizin	64 3.7	9 3.75	8.88	57 3.56	11 3.33	11.74	100 5.57	36 7.04	28.46
7 Agrar-/ Forst-/ Ernährungswissen-schaften	161 9.3	27 11.25	22.32	164 10.26	39 11.82	33.85	61 3.39	5 0.97	17.32
8 Ingenieur-wissenschafen	--	--	--	--	--	--	--	--	--
9 Kunst, Kunstwissenschaften	--	--	--	--	--	--	--	--	--
10 Lehrer	--	--	--	--	--	--	326 18.16	83 16.24	92.79
Gesamt	1726 100%	240 100%	240 100%	1597 100%	330 100%	330 100%	1795 100%	511 100%	511 100%
χ2-Wert df	7.56 n.s. 5			15.6** 5			147.16*** 6		
	1) 0.002 2) 0.14 3) 3.29 4) 3.15 5) 0.00 6) 0.98 --- df=5 7.56			1) 0.01 2) 2.02 3) 10.46 4) 2.29 5) 0.04 6) 0.78 --- df=5 15.6			1) 0.31 2) 102.2 3) 29.46 4) 3.41 5) 1.99 6) 8.76 7) 1.03 --- df=6 147.16		

* signifikant (p<=0.05) ** hoch signifikant (p<=0.01) *** höchst signifikant (p<=0.001)

Quelle: Berufsverlaufsstudie Ostdeutschland, eigene Berechnungen

Tab. 4.4: χ^2 –Anpassungstests für Hochschulabsolventen in Rostock

	Rostock 1985			Rostock 1990			Rostock 1995		
Fächergruppe	GG	Netto-samp.	erwartete Häufigkeit	GG	Netto-samp.	erwartete Häufigkeit	GG	Netto-samp.	erwartete Häufigkeit
1 Sprach-/ Kulturwissenschaften	237 18.03	55 20.3	48.86	205 16.77	59 19.73	50.14	24 1.9	15 2.83	10.7
2 Sport	54 4.1	4 1.48	11.11	35 2.86	2 0.67	8.55	--	--	--
3 Rechts-/ Wirtschafts-/ Sozialwissenschaften	111 8.4	8 2.95	22.76	55 4.5	10 3.34	13.45	122 10.11	50 9.43	53.58
4 Mathematik, Naturwissenschaften	135 10.27	32 11.81	27.83	122 9.9	34 11.37	29.6	89 7.3	53 10	38.69
5 Humanmediziner	279 21.23	52 19.19	57.53	298 24.38	63 21.07	72.89	201 16.66	82 15.47	88.29
6 Veterinärmedizin	--	--	--	--	--	--	--	--	--
7 Agrar-/ Forst-/ Ernährungswissen-schaften	251 19.1	60 22.14	51.76	226 18.49	65 21.74	55.28	64 5.3	18 3.4	28.09
8 Ingenieur-wissenschafen	247 18.79	60 22.14	50.92	281 22.99	66 22.07	68.74	427 35.4	197 37.7	187.62
9 Kunst, Kunstwissenschaften	--	--	--	-	--	--	--	--	--
10 Lehrer	--	--	--	--	--	--	279 23.13	117 22.07	122.58
Gesamt	1314 100%	271 100%	271 100%	1222 100%	299 100%	299 100%	1206 100%	530 100%	530 100%
χ2-Wert df	18.96** 6			11.24 n.s. 6			12.01 n.s. 6		
	1) 0.77 2) 4.55 3) 9.57 4) 0.62 5) 0.53 6) 1.31 7) 1.61 --- df=6 18.96			1) 1.56 2) 5.01 3) 0.88 4) 0.65 5) 1.34 6) 1.70 7) 0.10 --- df=6 11.24			1) 1.72 2) 0.23 3) 5.29 4) 0.44 5) 3.62 6) 0.46 7) 0.25 --- df=6 12.01		

* signifikant (p<=0.05) ** hoch signifikant (p<=0.01) *** höchst signifikant (p<=0.001)

Quelle: Berufsverlaufsstudie Ostdeutschland, eigene Berechnungen

4.7 Originaltabellen ohne Zusammenfassungen

Tab. 4.5: Ursprungstabelle der Hochschulabsolventen

	Rostock 1985		Rostock 1990		Rostock 1995		Leipzig 1985		Leipzig 1990		Leipzig 1995	
Fächergruppe	GG	Netto-sample	GG	Netto-sample	GG	Netto-sample	GG	Netto-sample	GG	Netto-sample	GG	Netto-sample
1 Sprach-/ Kulturwissen-schaften	237 18%	55 20.3	205 17%	59 19.73	24 2%	15 2.8%	529 31%	75 31.25	533 33%	107 32.42	186 10.4%	57 11.15
2 Sport	54 4%	4 1.48	35 3%	2 0.67	0 0	1 0.18%	--	--	--	--	88 4.9%	2 0.39
3 Rechts-/ Wirtschafts-/ Sozialwissen-schaften	111 8%	8 2.95	55 4%	10 3.34	122 10.1%	50 9.43%	298 17%	39 16.25	237 15%	39 11.82	287 16%	173 33.86
4 Mathematik, Naturwissen-schaften	135 10%	32 11.81	122 10%	34 11.37	89 7.4%	53 10%	220 13%	42 17.5	199 12%	64 19.39	447 24.9%	60 11.74
5 Humanmedizi-ner	279 21%	52 19.19	298 24%	63 21.07	201 16.7%	82 15.47	397 23%	42 17.5	369 23%	63 19.09	388 21.6%	91 17.81
6 Veterinärmedi-zin	--	--	--	--	--	--	64 4%	9 3.75	57 4%	11 3.33	100 5.6%	36 7.05
7 Agrar-/ Forst-/ Ernährungswis-senschaften	251 19%	60 22.14	226 18%	65 21.74	64 5.3%	18 3.4	161 9%	27 11.25	164 10%	39 11.82	61 3.4%	5 0.98
8 Ingenieur-wissenschafen	247 19%	60 22.14	281 23%	66 22.07	427 35.4%	197 37.7	21 1%	2 0.83	20 1%	3 0.91	--	6 1.17
9 Kunst, Kunstwissen-schaften	--	--	--	--	11 0.91%		36 2%	4 1.67	18 1%	4 1.21	11 0.6%	2 0.39
10 Lehrer	--	--	--	--	257 21.31%	115 21.16	--	--	--	--	227 12.6%	79 15.45
Gesamt	1314 100%	271 100	1222 100%	299 100	1206 100%	530 100	1726 100%	240 100	1597 100%	330 100	1795 100%	511 100

Tab. 4.6.: Ursprungstabelle der Lehrabsolventen

	Rostock 1985		Rostock 1990		Rostock 1995		Leipzig 1985		Leipzig 1990		Leipzig 1995	
Fachgruppen	GG	Netto-sample	GG	Netto-sample	GG	Netto-sample	GG	Netto-sample	GG	Netto-sample	GG	Netto-sample
1.Pflanzenbauer/ Tierzüchter/ Fischereiberufe	81 1%	--	100 2%	--	--	--	53 1%	--	80 1%	--	--	--
2.Bergleute/ Mineralgewinner	--	--	--	--	--	--	--	--	--	--	--	--
3.Fertigungs-berufe	3100 56%	107 53.5%	2185 49%	116 49.15%	696 42.3%	115 42.28%	3748 67%	109 57.98%	3437 61%	159 63.34%	1582 48.3%	131 42.67%
4.Technische Berufe	89 2%	4 2%	55 1%	4 1.69%	31 1.9%	6 2.2%	218 4%	11 5.85%	187 3%	4 1.59%	22 0.7%	3 0.97%
5.Dienstleistungs-berufe	2290 41%	89 44.5%	2154 48%	116 49.15%	919 55.8%	151 55.51%	1579 28%	68 36.17%	1917 34%	88 35.06%	1670 51%	173 56.35%
Gesamt	5560 100%	200 100%	4494 100%	236 100%	1646 100%	272 100%	5589 100%	188 100%	5621 100%	251 100%	3274 100%	307 100%

4.8 Originaltabellen der ersten Welle ohne Nachkodierungen

Abb. 4.7: Vergleich Stichprobenziehung der Hochschulabsolventen mit Responseraten

	Rostock 1985		**Rostock 1990**		**Leipzig 1985**		**Leipzig 1990**	
Fächergruppe	**Grundge-samtheit**	**Rücklauf**	**Grundge-samtheit**	**Rücklauf**	**Grundge-samtheit**	**Rücklauf**	**Grundge-samtheit**	**Rücklauf**
1 Sprach-/ Kulturwissen-schaften	237 (18 %)	54 (20.1 %) <2.1>	205 (17 %)	59 (19.9 %) <1.9>	529 (31 %)	74 (31.0%) <0>	533 (33 %)	105 (32.5 %) <0.5>
2 Sport	54 (4 %)	4 (1.5 %) <2.5>	35 (3 %)	2 (0.7 %) <2.3>	-	-	-	-
3 Rechts-/ Wirtschafts-/ Sozialwissenschaften	111 (8 %)	8 (3 %) <5>	55 (4 %)	10 (3.4 %) <0.6>	298 (17 %)	39 (16.3%) <0.7>	237 (15 %)	38 (11.8 %) <3.2>
4 Mathematik, Natur-wissenschaften	135 (10 %)	32 (11.9 %) <1.9>	122 (10 %)	34 (11.4 %) <1.4>	220 (13 %)	42 (17.6%) <4.6>	199 (12 %)	63 (19.5 %) <7.5>
5 Humanmedizin	279 (21 %)	52 (19.4 %) <1.6>	298 (24 %)	63 (21.2 %) <2.8>	397 (23 %)	42 (17.6%) <5.4>	369 (23 %)	62 (19.2%) <3.8>
6 Veterinärmedizin	-	-	-	-	64 (4 %)	9 (3.8 %) <0.2>	57 (4 %)	11 (3.4 %) <0.6>
7 Agrar-/ Forst-/ Ernäh-rungswissenschaften	251 (19 %)	59 (22.0%) <3>	226 (18 %)	64 (21.5 %) <3.5>	161 (9 %)	27 (11.3%) <2.3>	164 (10 %)	37 (11.5%) <1.5>
8 Ingenieurwissenschaf-ten	247 (19 %)	59 (22.0 %) <3>	281 (23 %)	65 (21.9 %) <1.1>	21 (1 %)	2 (0.8 %) <0.2>	20 (1 %)	3 (0.9 %) <0.1>
9 Kunst, Kunstwissen-schaften	-		-	-	36 (2 %)	4 (1.7 %) <0.3>	18 (1 %)	4 (1.2 %) <0.2>
Gesamt	1314 (100%)	268 (100%)	1222 (100%)	297 (100 %)	1726 (100%)	239 (100%)	1597 (100%)	323 (100%)

Abb. 4.8: Vergleich Stichprobenziehung und Responseraten der Lehrabsolventen [abs. - in %]

	Rostock 1985			**Rostock 1990**			**Leipzig 1985**			**Leipzig 1990**		
Fachgruppen	Grund-gesamt-heit	Brutto-sample	Rück-lauf-quoten	Grund-gesamt-heit	Brutto-sample	Rück-lauf-quoten	Grund-gesamt-heit	Brutto-sample	Rück-lauf-quoten	Grund-gesamt-heit	Brutto-sample	Rück-lauf-quoten
1 Pflanzenbauer/ Tierzüchter/ Fischereiberufe	81 (1 %)	-	-	100 (2 %)	-	-	53 (1 %)	-	-	80 (1 %)	-	-
2 Bergleute/ Mineralgewinner	-	-		-	-		-	-		-	-	-
3 Fertigungsberufe	3100 (56)	566 (57)	107 (53.5) <3.5>	2185 (49)	497 (50)	106 (44.9) <5.1>	3748 (67)	677 (68)	109 (58.0) <10>	3437 (61)	637 (64)	143 (57.2) <6.6>
4 Technische Berufe	89 (2)	16 (2)	4 (2) <0>	55 (1)	11 (1)	4 (1.7) <0.7>	218 (4)	39 (4)	11 (5.9) <1.9>	187 (3)	9 (1)	4 (1.6) <0.6>
5 Dienstleistungsberufe	2290 (41)	418 (42)	89 (44.5) <2.5>	2154 (48)	490 (49)	126 (53.4) <4.4>	1579 (28)	284 (28)	68 (36.2) <8.2>	1917 (34)	354 (35)	103 (41.0) <6>
Gesamt	5560 (100)	1000 (100)	200 (100)	4494 (100)	998 (100)	236 (100)	5589 (100)	1000 (100)	188 (100)	5621 (100)	1000 (100)	250 (100)

5 Literatur

Bedau, Klaus Dietrich, Heinz Vortmann (1991): Vergleich der DDR- Basisbefragung '90 des sozio-ökonomischen Panels mit makro- orientierten Rahmendaten für das Jahr 1988, in: Projektgruppe „Das sozio- ökonomische Panel" (Hrsg.) (1991).

Blossfeld, H.-P./Rohwer, G. (1995): Techniques of Event History Modeling. New Approaches to Causal Analysis. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Blossfeld, Hans-Peter (1989): Zur Repräsentativität der SFB-3-Lebensverlaufsstudie – Ein Vergleich mit Daten der amtlichen Statistik, in: Karl Ulrich Mayer, Erik Brückner (Hrsg.): Lebensverläufe und Wohlfahrtsentwicklung. Konzeption, Design und Methodik der Erhebung von Lebensläufen der Geburtsjahrgänge 1929-31, 1939-41, 1949-51, Teil I. Materialien aus der Bildungsforschung Nr. 35, Berlin.

Böltken, Ferdinand (1976): Auswahlverfahren. Stuttgart: Teubner.

Brückner, E. (1990): Die retrospektive Erhebung von Lebensverläufen. In: Mayer, K. U. (Hg.): Lebensverläufe und sozialer Wandel, [KZfSS], Sonderheft, S. 374-403

Bundesforschungsanstalt für Landeskunde und Raumordnung (Hrsg.) (1996): Regionale Schwerpunkte der Arbeitslosigkeit in Deutschland, in: Mitteilungen und Informationen (BfLR), Nr. 1/96, Ausgabe Februar

Diekmann, Andreas (1998): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen, Reinbek b. Hamburg: Rohwolt.

Dillman, Don A. (1983): Mail and Other Self-Administered Questionnaires. In P. H. Rossi u.a. (Hrsg.): Handbook of survey research. New York: Academic Press. S. 359-377.

Frey, James H./ Kunz, Gerhard/ Lüschen, Günther (1990): Telefonumfragen in der Sozialforschung. Opladen: Westdeutscher.

Graff, Jörg (1992): Soziologische Statistik, Pfaffenweiler: Centaurus.

Hartmann, Peter H., Bernhard Schimpl-Neimanns (1992): Sind Sozialstrukturanalysen mit Umfragedaten möglich? Analysen zur Repräsentativität einer Sozialforschungsumfrage, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie Jg. 44, H 2, 1992, S. 315- 340.

Hartmann, Peter H., Bernhard Schimpl-Neimanns (1993): Affirmative Repräsentativitäts „beweise". Test konkreter Verteilungsabweichungen?, in: Bortz, Jürgen (1989): Statistik. Für Sozialwissenschaftler, Berlin: Springer.

Hippler, Hans-J. (1988): Methodische Aspekte schriftlicher Befragungen: Probleme und Forschungsperspektiven. In planung und analyse 15: 244-248.

Kreienbrock, Lothar (1989): Einführung in die Stichprobenverfahren. München: Oldenbourg.

Kromrey, Helmut (1998): Empirische Sozialforschung, Opladen: Leske + Budrich.

Mayer, Karl Ulrich/ Brückner, Erika (1989): Lebensverläufe und Wohlfahrtsentwicklung. Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1929-1931, 1939-1941, 1949-1951. Materialien aus der Bildungsforschung 35. Zwei Teile. Berlin: MPI.

Merten, K.; Teipen, P. (1991): Empirische Kommunikationsforschung. München.

Minks, Karl-Heinz/ Bathke, Gustav-Wilhelm (1992): Berufliche Integration und Weiterbildung von jungen Akademikern aus den neuen Bundesländern. Hannover: HIS.

Porst, Rolf (1999): Thematik oder Incentives? Zur Erhöhung der Rücklaufquoten bei postalischen Befragungen, in: ZUMA Nachrichten 45, 1999, S. 72 –87.

Prein, Gerald, Susann Kluge, Udo Kelle (1994): Strategien zur Sicherung von Repräsentativität und Stichprobenvalidität bei kleinen Samples. Arbeitspapier Nr. 18, des Sonderforschungsbereichs 186 an der Universität Bremen.

Projektgruppe „Das sozio- ökonomische Panel“ (Hrsg.) (1991): Lebenslagen im Wandel: Basisdaten und –analyse zur Entwicklung in den neuen Bundesländern, Frankfurt/M: Campus.

Reuband, Karl- Heinz (1999): Postalische Befragungen in den neuen Bundesländern: Durchführungsbedingungen, Ausschöpfungsquoten und soziale Zusammensetzung der Befragten in einer Großstadtstudie, in: ZA- Information 45, 1999, S. 71-99.

Rendtel, Ulrich (1995): Lebenslagen im Wandel: Panelausfälle und Panelrepräsentivität, New York: Campus.

Rendtel, Ulrich, Ulrich Pötter (1992): Über den Sinn und Unsinn von Repräsentativitätsstudien, DIW Diskussionspapier Nr. 61.

Rendtel, Ulrich, Ulrich Pötter (1993): „Empirie“ ohne Daten. Kritische Anmerkungen zu einer Repräsentativitätsstudie über den Allbus, in: Kölner Zeitschrift für Soziologie und Sozialpsychologie Jg. 45, H 2, 1993, S. 350- 358.

Rohwer, G. (1992a): Ein R/ZOO Archiv für die Daten des kumulierten ALLBUS, 1980-1988. In: ZA-Information 31: 100-106.

Rohwer, G. (1992b): A Retrieval Program for the SOEP. Version 1.0. Universität Bremen (EMPAS).

Rohwer, G. (1994a): Ein R/ZOO. Ein Retrievalprogramm für das Sozio-ökonomische Panel. Version 94.1. Berlin: DIW.

Rohwer, G. (1994b): TDA Working Papers, 0-1 bis 9-1. Version 2. Universität Bremen (EMPAS).

Rohwer, Götz (1994): Kontingente Lebensverläufe. Soziologische und statistische Aspekte ihrer Beschreibung und Erklärung, Bremen.

Sahner, Heinz (1990): Schließende Statistik, Stuttgart: Teubener.

Sahner, Heinz (1991): Schließende Statistik, Stuttgart: Teubener.

Schnell, R./Hill, P. B./Esser, E. (1989): Methoden der empirischen Sozialforschung. 2. überarbeitete und erweiterte Aufl. München/Wien.

Schnell, Rainer, Paul B. Hill, Elke Esser (1993): Methoden der empirischen Sozialforschung, München: Oldenbourg.

Schober, K. (1993): Veränderte Statuspassagen und Strukturwandel in der beruflichen Erstausbildung. Ergebnisse einer Längsschnittuntersuchung bei Jugendlichen in den neuen Ländern, MittAB 2/93, S. 162-175

Statistisches Bundesamt (Hrsg.)(1992): Bildung und Kultur. Fachserie 11. Reihe 3. Berufliche Bildung. Stuttgart: Metzler-Poeschel.

Statistisches Bundesamt (Hrsg.)(1993): Bildung und Kultur. Fachserie 11. Reihe 4.1. Studenten an Hochschulen. Sommersemester 1992. Stuttgart: Metzler-Poeschel.

Teichler, Ulrich/ Winkler, Helmut (1990): Der Berufsstart von Hochschulabsolventen. Studien zu Bildung und Wissenschaft 87. Bonn: BMBW.

Windzio, Michael (2000): Transformation als Experiment sozialen Wandels. Die Beobachtung von Lebensverläufen, in: Reinhold Sackmann, Ansgar Weymann, Matthias Wingens (2000) (Hrsg.): Die Generation der Wende. Arbeitsmarkt, Familie und Beruf, im Erscheinen.

6 Anhang

Fragebogen der Berufsverlaufsstudie Ostdeutschland, Welle 2, Kohorten 1985 und 1990.