

**Forschungsdatenzentrum
der Statistischen
Landesämter**



Amtliche Mikrodaten für die Sozial- und Wirtschaftswissenschaften

**Beiträge zu den Nutzerkonferenzen
des FDZ der Statistischen Landesämter 2005**

Impressum

Herausgeber: Forschungsdatenzentrum der Statistischen Landesämter
Internet: www.forschungsdatenzentrum.de
E-Mail: forschungsdatenzentrum@lds.nrw.de

Gesamtherstellung: Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen
Mauerstraße 51, 40476 Düsseldorf
Postfach 10 11 05, 40002 Düsseldorf
Telefon 0211 9449-01 • Telefax 0211 442006
Internet: www.lds.nrw.de
E-Mail: poststelle@lds.nrw.de

Für die Inhalte der einzelnen Fachbeiträge sind die jeweiligen Autoren verantwortlich.

Erschienen im Mai 2006

© Statistische Ämter der Länder

Für nicht gewerbliche Zwecke sind Vervielfältigung und unentgeltliche Verbreitung, auch auszugsweise, mit Quellenangabe gestattet. Die Verbreitung, auch auszugsweise, über elektronische Systeme/Datenträger bedarf der vorherigen Zustimmung. Alle übrigen Rechte bleiben vorbehalten.

Bestell-Nr. S 61 3 2005 51
ISBN 3-935372-87-6

Vorwort

Im Frühjahr 2005 führte das Forschungsdatenzentrum der Statistischen Landesämter in Kooperation mit Universitäten und wissenschaftlichen Einrichtungen bundesweit vier regionale Nutzerkonferenzen durch. Wissenschaftlerinnen und Wissenschaftler stellten hier ihre aktuellen Forschungsprojekte auf der Basis amtlicher Mikrodaten vor. Mit insgesamt etwa 30 Fachvorträgen und über 300 Teilnehmern unterstreichen die Veranstaltungen in Berlin, Kiel, Köln und Stuttgart das zunehmende Interesse der Wissenschaft an der Analyse von Daten aus der amtlichen Statistik. Das thematische Spektrum der Vorträge, das sich über die Bereiche Wirtschaft, Soziales, Agrikultur und Umwelt erstreckt, belegt zudem die Eignung amtlicher Mikrodaten für die Bearbeitung innovativer Fragestellungen aus den unterschiedlichsten wissenschaftlichen Disziplinen.

Die vorliegende Veröffentlichung umfasst eine Auswahl der Beiträge zu Forschungsprojekten auf Basis der Sozial- und Wirtschaftsstatistiken. Eine Zusammenstellung der Beiträge zu agrar- und umweltwissenschaftlichen Themen findet sich in einem weiteren, zeitgleich erscheinenden Band.

Unser Dank gilt den Referentinnen und Referenten, die mit ihren Beiträgen anschaulich das Analysepotenzial der Mikrodaten aus der amtlichen Statistik aufzeigen. Ebenso möchten wir uns bei Professor Dr. Hans-Jürgen Andreß, Universität zu Köln, und Professor Dr. Ulrich Rendtel, Freie Universität Berlin, für ihre Unterstützung bei der Ausrichtung der beiden Konferenzen zu sozial- und wirtschaftswissenschaftlichen Themen bedanken. Das Engagement und die Gastfreundschaft unserer Kooperationspartner und deren Mitarbeiterinnen und Mitarbeiter haben maßgeblich zum Erfolg der Veranstaltungen beigetragen.

Die in den beiden Bänden vorgestellten Forschungsprojekte stehen als Beispiele für den Erfolg eines kontinuierlichen Ausbaus der informationellen Infrastruktur zwischen Wissenschaft und amtlicher Statistik. Sie sollten ebenso als Aufforderung an Forscherinnen und Forscher verstanden werden, das ständig wachsende Datenangebot des Forschungsdatenzentrums der Statistischen Landesämter intensiv für ihre Projekte zu nutzen. Wir sind überzeugt, dass die folgenden Beiträge gute Ideen und Anregungen für weitere Forschung auf Basis amtlicher Daten liefern können. Für das Forschungsdatenzentrum der Statistischen Landesämter

Jochen Kehlenbach
*Präsident des Landesamtes für
Datenverarbeitung und Statistik
Nordrhein-Westfalen*

Prof. Dr. Ulrike Rockmann
*Präsidentin des
Statistischen Landesamtes
Berlin*

Amtliche Mikrodaten für die Sozial- und Wirtschaftswissenschaften

Beiträge zu den Nutzerkonferenzen des FDZ der Statistischen Landesämter 2005

	Seite
Datenangebot und Datenzugang im Forschungsdatenzentrum der Statistischen Landesämter	
Sylvia Zühlke, Helga Christians	7
The Effects of Public Schooling Expenditures on Individual Labour Market Outcomes	
Andreas Ammermüller, Anja Kuckulenz, Thomas Zwick	23
Kein Bedarf an mehr Akademikerinnen und Akademikern	
Christiane Mück, Karen Mühlenbein	39
Möglichkeiten der Analyse von Arbeitsmarktübergängen mit Daten des Mikrozensus-Panels	
Michael Konold	55
Analysen regionalspezifischer Verhaltensmuster der Wohnungsnachfrage	
Irene Iwanow	69
Möglichkeiten kleinräumiger Analysen auf Basis des Mikrozensus	
Helga Christians	81
Small Area Estimation beim Zensus 2011	
Kersten Magg, Ralf Münnich, Josef Schäfer	93
Zum Einfluss von Marktfaktoren, institutionellen und individuellen Faktoren auf Jugendarbeitslosigkeit in Westeuropa	
Katrin Golsch	107
Möglichkeiten und Grenzen der Nutzung der Einzeldaten der amtlichen Kinder- und Jugendhilfestatistik	
Matthias Schilling	125

	Seite
Analyse privater Vermögen auf der Basis der Einkommens- und Verbrauchsstichproben	
Peter Westerheide	133
Krankenhaus-Diagnosestatistik – Scientific-Use-File und beispielhafte Analysen	
Gabriele Philipp, Olaf Schoffer	147
Tarifbindung, betriebliche Lohnhöhe und Lohnstreuung im Produzierenden Gewerbe Baden-Württembergs	
Stephan Bechtel, Wolf Dieter Heinbach, Harald Strotmann	159
Erhebungen zu Verdiensten und Arbeitskosten	
Hans-Peter Hafner	179
Erste Scientific-Use-Files aus den Wirtschaftsstatistiken	
Roland Sturm, Rainer Lenz	191
Qualitätssicherung des Faktors Interviewer für den Mikrozensus	
Christin Schäfer, Hartmut Bömermann, Ricarda Nauenburg, Karsten Wenzel, Klaus-Robert Müller, Gert G. Wagner,	209
Datenangebot	219

Datenangebot und Datenzugang im Forschungsdatenzentrum der Statistischen Landesämter

Sylvia Zühlke, Helga Christians

Das Forschungsdatenzentrum der Statistischen Landesämter wurde im Jahr 2002 als Serviceeinrichtung für die Wissenschaft gegründet. Die wesentlichen Ziele dieser Einrichtung bestehen darin, gemeinsam mit dem Forschungsdatenzentrum des Statistischen Bundesamtes die Dateninfrastruktur zu verbessern und den Zugang zu Einzeldaten der amtlichen Statistik zu erleichtern (siehe Zühlke/Zwick/Schornhorst/Wende 2005). Nachdem in der Anfangszeit der Aufbau der Infrastruktur im Vordergrund stand, wird das neue Dienstleistungsangebot nunmehr intensiv genutzt. Der folgende Beitrag beschreibt zunächst die Aufgaben des Forschungsdatenzentrums der Statistischen Landesämter und stellt im Anschluss daran die Zugangswege sowie das Datenangebot vor.

1 Aufgaben des Forschungsdatenzentrums

Das Forschungsdatenzentrum ist eine Arbeitsgemeinschaft aller statistischen Landesämter. Es zeichnet sich mit seinen 16 regionalen Standorten insbesondere durch eine starke bundesweite Präsenz aus. Dadurch besteht für Wissenschaftler/-innen die Möglichkeit, in der Nähe des eigenen Arbeitsplatzes Daten der amtlichen Statistik zu nutzen und das Beratungsangebot des Forschungsdatenzentrums in Anspruch zu nehmen. Die regionalen Standorte haben außerdem fachliche Schwerpunkte, im Rahmen derer sie ausgewählte Statistiken bearbeiten. Hierdurch werden ein breites Datenangebot und die fundierte fachliche Beratung der Nutzerinnen und Nutzer sichergestellt. Unabhängig von den fachlichen Schwerpunkten der regionalen Standorte können die Nutzer/-innen in dem jeweiligen Standort Zugang zu dem gesamten Datenangebot des Forschungsdatenzentrums erhalten.

Im Einzelnen nimmt das Forschungsdatenzentrum der Statistischen Landesämter folgende Aufgaben wahr:

- 1) Eine wesentliche Aufgabe des Forschungsdatenzentrums besteht darin, die Dateninfrastruktur für die Politikberatung und die wissenschaftliche Forschung zu verbessern. Um dieses Ziel zu erreichen, wurden zwei Schwerpunkte definiert: Zunächst müssen die Mikrodaten der amtlichen Statistik für die einzelnen Bundesländer an einer Stelle zusammengeführt werden. Dies ist notwendig, da die Einzeldaten aufgrund des föderalen Aufbaus der amtlichen Statistik in der Regel dezentral gespeichert werden. Das heißt, dass jedes statistische Amt meist nur über die Einzeldaten verfügt, die es für sein Bundesland erhoben hat. Da sich wissenschaftliche Analysen aber häufig auf mehrere Bundesländer beziehen, ist eine Zusammenführung der Daten an einer zentralen Stelle eine wesentliche Voraussetzung für die Verbesserung der informationellen Infrastruktur. Als zweiter Schwerpunkt ist die Dokumentation der zusammengeführten Mikrodaten durch umfangreiche Metadaten zu nennen. Die Metadaten umfassen nicht nur eine technische Beschreibung der Datensatzstrukturen, sondern sie enthalten auch alle methodischen Informationen über die Erhebung und Aufbereitung der Daten sowie deren Qualität. Ohne diese Angaben ist eine sinnvolle Interpretation der Analyseergebnisse nur schwer möglich. Die Metadaten sollen den Wissenschaftler/-innen im Rahmen eines Informationssystems zukünftig über das Internet zur Verfügung gestellt werden.

- 2) Die zweite Aufgabe besteht darin, den Zugang der Wissenschaft zu den Einzeldaten der amtlichen Statistik durch die Einrichtung verschiedener Zugangswege zu erleichtern. Den Datennutzern stehen drei unterschiedliche Möglichkeiten zur Verfügung. Dies sind die Scientific-Use-Files zur Off-Site-Nutzung einerseits und die Gastwissenschaftlerarbeitsplätze sowie die kontrollierte Datenfernverarbeitung zur On-Site-Nutzung andererseits. Diese Nutzungsmöglichkeiten werden weiter unten näher beschrieben.
- 3) Als dritte Aufgabe ist die Beratung der Datennutzer/-innen zu nennen. Diese umfasst zum einen die allgemeine Beratung zu den Nutzungswegen und zum Datenangebot des Forschungsdatenzentrums. Zum anderen sind die einzelnen regionalen Standorte auf ausgewählte fachliche Schwerpunkte spezialisiert, sodass hierdurch für die Datennutzer/-innen auch eine intensive fachliche Beratung sichergestellt wird.
- 4) Die vierte Aufgabe des Forschungsdatenzentrums besteht darin, gemeinsam mit der Wissenschaft inhaltliche und methodische Forschungsprojekte durchzuführen. Dabei geht es vor allem um die Verbesserung des Analysepotenzials sowie um die Entwicklung neuer statistischer Methoden.

Seit der Einrichtung des Forschungsdatenzentrums wurden in den einzelnen Aufgabenbereichen große Fortschritte erzielt: So konnte das Mikrodatenangebot für die Wissenschaft deutlich erweitert werden, für alle Statistiken stehen Basisinformationen und für ausgewählte Statistiken umfangreiche Metadaten zur Verfügung. Das neue Dienstleistungsangebot kann in allen regionalen Standorten genutzt werden. Außerdem wurde die Wissenschaft im Rahmen einer breit angelegten Öffentlichkeitsarbeit über die Möglichkeiten, die das Forschungsdatenzentrum bietet, informiert. Mittlerweile wird das Angebot an amtlichen Mikrodaten intensiv für Forschungsvorhaben genutzt.

2 Nutzungswege

Die amtliche Statistik darf Einzelangaben zur Verfügung stellen, sofern diese absolut anonymisiert sind. Um die absolute Anonymität sicherzustellen, werden die Daten z. B. durch Aggregation oder durch die Entfernung einzelner Merkmale so weit verändert, dass eine Identifizierung der Auskunftgebenden nach menschlichem Ermessen unmöglich gemacht wird. Die amtliche Statistik bietet absolut anonymisierte Mikrodaten in Form so genannter Public-Use-Files an. Eine weitere Hauptzielrichtung der Public-Use-Files liegt im Bereich der Hochschullehre. Die Forschungsdatenzentren entwickeln so genannte Campus-Files, die an Hochschulen zu Lehrzwecken eingesetzt werden können. Diese Datensätze sollen Student/-innen frühzeitig die Möglichkeit bieten, die Besonderheiten der Analyse von amtlichen Mikrodaten kennen zu lernen.

Absolut anonymisierte Daten weisen in Folge der Anonymisierung einen hohen Verlust an Informationen auf. Sie eignen sich in der Regel nicht für die Analyse differenzierter Forschungsfragen. Die Forschungsdatenzentren haben daher für die Wissenschaft die drei bereits genannten Nutzungswege eingerichtet. Sie leiten sich aus den rechtlichen Rahmenbedingungen ab, unter denen die amtliche Statistik der Wissenschaft Zugang zu Einzeldaten gewähren darf. Sowohl die informationelle Selbstbestimmung als auch die Wissenschaftsfreiheit sind als Grundrechte in der Verfassung verankert. Damit ist der Gesetzgeber aufgerufen, für einen angemessenen Ausgleich dieser Grundrechte Sorge zu tragen. Bei der Novellierung des Bundesstatistikgesetzes im Jahr 1987

wurde dies berücksichtigt, indem in § 16 Bundesstatistikgesetz (BStatG) festgeschrieben wurde, dass der Wissenschaft Daten übermittelt werden können, die eine Deanonymisierung zwar nicht mit absoluter Sicherheit ausschließen, aber Betroffenen nur dann zugeordnet werden können, wenn der Datenempfänger einen unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft erbringt. Dieses so genannte „Wissenschaftsprivileg“ ist Voraussetzung für die Nutzung der als faktisch anonym bezeichneten Mikrodaten, die für ein definiertes Forschungsvorhaben bereitgestellt werden.

Für Wissenschaftler/-innen ist die Auswertung von Einzelangaben am eigenen Arbeitsplatz die bequemste Lösung. Daher wird eine solche Nutzungsmöglichkeit von der Wissenschaft eindeutig präferiert (siehe Zühlke/Hetke 2003). Die Forschungsdatenzentren entwickeln aus diesem Grund so genannte Scientific-Use-Files für die Off-Site-Nutzung, die an wissenschaftliche Einrichtungen übermittelt werden dürfen. Es handelt sich dabei um standardisierte Produkte, die einen festgelegten Merkmalsumfang haben. Solche Produkte liegen bereits für eine Reihe von haushalts- und personenbezogenen Statistiken vor. Auch für wirtschaftsstatistische Daten wurden schon erste Datenbestände erstellt (siehe Ronning/Gnoss 2003, Sturm/Lenz in diesem Band).

Um die Anonymität der Daten sicherzustellen, müssen bei den Scientific-Use-Files zur Off-Site-Nutzung Vergrößerungen bei der regionalen und/oder bei der fachlichen Gliederung vorgenommen werden. Die Erstellung von Scientific-Use-Files ist daher sehr aufwändig und stets mit einem – zum Teil nicht unerheblichen – Informationsverlust verbunden. Für viele wissenschaftliche Fragestellungen können sie daher unzureichend sein.

Aus diesem Grund werden zwei Wege zur On-Site-Nutzung angeboten, und zwar die Gastwissenschaftlerarbeitsplätze sowie die kontrollierte Datenfernverarbeitung. Die Gastwissenschaftlerarbeitsplätze befinden sich in den geschützten Räumen der statistischen Ämter. Aufgrund der geltenden Sicherheitsvorkehrungen dürfen die dort bereitgestellten Datenbestände mehr Informationen enthalten als die standardisierten Scientific-Use-Files. Die Ausgabe der Ergebnisse können die Wissenschaftler/-innen nur in dem statistischen Amt veranlassen, in dem sie gerade arbeiten. Dort werden die Auswertungen – vor einer Freigabe – auf Geheimhaltung geprüft. Zusätzlich wird im Vorfeld ein Nutzungsvertrag zwischen dem statistischen Amt und den Nutzer(inne)n geschlossen und die Wissenschaftler/-innen werden auf das Statistikgeheimnis verpflichtet. Beim Arbeiten mit den Daten am Gastwissenschaftlerarbeitsplatz können die Wissenschaftler/-innen auf die Unterstützung des Fachpersonals der statistischen Ämter zurückgreifen.

Der zweite Weg, Mikrodaten On-Site auszuwerten, besteht darin, eine kontrollierte Datenfernverarbeitung durchzuführen. Hierfür erstellen die Wissenschaftler/-innen zu ihrem Forschungsvorhaben ein Auswertungsprogramm. Der durchführende Standort des Forschungsdatenzentrums prüft das Programm und wendet es auf formal anonymisierte Mikrodaten an.¹⁾ Die Ergebnisse werden vor der Weitergabe an die Wissenschaftler/-innen auf Wahrung der Geheimhaltung geprüft. Die Wissenschaftler/-innen haben also keinen direkten Kontakt mit den geheimhaltungsbedürftigen Mikrodaten. Eine weitergehende Anonymisierung der für die Auswertungen genutzten Mikrodaten ist daher nicht notwendig.

Auf Grundlage der beschriebenen Zugangswege kann im Forschungsdatenzentrum ein vielfältiges Datenangebot genutzt werden. Dieses wird im folgenden Abschnitt dargestellt.

1) Formal anonymisierte Daten enthalten in der Regel keine direkten Identifikatoren wie den Namen oder die Adresse der Erhebungseinheit. Die Erhebungsmerkmale sind jedoch in vollem Umfang in den Daten enthalten.

3 Datenangebot

Während vor der Einrichtung des Forschungsdatenzentrums der Statistischen Landesämter lediglich acht unterschiedliche Erhebungen der amtlichen Statistik für wissenschaftliche Analysen zugänglich waren, besteht das Datenangebot zurzeit aus über 40 Statistiken mit verschiedenen fachlichen Schwerpunkten. So werden neben haushalts- und personenbezogenen Datensätzen auch Informationen aus den Bereichen Wirtschaft, Agrar, Umwelt, Recht und Steuern bereit gestellt.²⁾ Einen Überblick über die zurzeit in den jeweiligen fachlichen Bereichen verfügbaren Statistiken findet sich im letzten Teil des Bandes. Das Datenangebot wird ständig erweitert. Informationen über das aktuelle Datenangebot sind unter www.forschungsdatenzentrum.de verfügbar.

Im Folgenden werden einige Statistiken genauer dargestellt. Für diesen Überblick wurden Datenbestände ausgewählt, die sich insbesondere für drei zentrale Analysebereiche eignen, und zwar für Untersuchungen

- des Arbeitsmarktes, des Bildungssystems und der Familiensituation,
- der Einkommenssituation und
- der betrieblichen Situation.

Die vorliegende Dokumentation beinhaltet weiterhin eine Darstellung der amtlichen Mikrodaten aus dem Bereich der Gesundheitsstatistik (siehe Philipp/Schoffer in diesem Band). Einen Überblick über die agrar- und umweltwissenschaftlichen Analysen auf der Basis amtlicher Mikrodaten enthält die parallel erscheinende Dokumentation zu den entsprechenden Nutzerkonferenzen des Forschungsdatenzentrums der Statistischen Landesämter (2006).³⁾

3.1 Mikrodaten zur Analyse des Arbeitsmarktes, des Bildungssystems und der Familiensituation

Als Mehrthemenbefragung ist der **Mikrozensus** die zentrale Datenquelle für Analysen zum Arbeitsmarkt, zum Bildungssystem sowie zu familiären Prozessen. Beim Mikrozensus handelt es sich um eine 1 %-Haushaltsstichprobe, bei der etwa 370 000 Haushalte und 820 000 Personen befragt werden. Das Mikrozensus-Frageprogramm besteht aus einem Grund-, einem Ergänzungs- und einem Zusatzprogramm. Das Grundprogramm wird jährlich mit einem Auswahlsatz von 1 % erhoben. Hier werden Fragen zu Merkmalen der Person, zur Haupt- und Nebenwohnung, zur Staatsangehörigkeit, zum Familien- und Haushaltszusammenhang, zum Erwerbsverhalten, zur Schul- und Berufsausbildung, zu den Quellen des Lebensunterhalts, zur Höhe des Einkommens sowie zur Kranken- und Rentenversicherung gestellt. Das Ergänzungsprogramm wird ebenfalls jährlich erhoben, allerdings nur mit einem Auswahlsatz bis zu 0,5 %. Es umfasst Informationen zur beruflichen und allgemeinen Aus- und Fortbildung, zum Wohnsitz ein Jahr vor

2) Die Entwicklung des Datenangebotes basiert auf den Ergebnissen einer Nutzerbefragung, die im Jahr 2002 durchgeführt wurde, sowie den Empfehlungen des Gründungsausschusses des Rates für Sozial- und Wirtschaftsdaten (siehe Zühlke/Hetke 2003). Weiterhin wurde der darüber hinaus gehende aktuelle Datenbedarf der Wissenschaft, der in Form von Nutzungsanfragen an das Forschungsdatenzentrum herangetragen wurde, bei der Entwicklung des Datenangebotes berücksichtigt. – 3) Weitere Artikel zu Vorträgen auf den Nutzerkonferenzen sind in Schmollers Jahrbuch, 2005, Jg. 125, Heft 4 erschienen.

der Erhebung, zur Pflegebedürftigkeit und Pflegeversicherung. Weiterhin werden ergänzende Angaben zur aktuellen und früheren Erwerbstätigkeit und zur Erwerbsbeteiligung erfragt. Die Zusatzprogramme werden dagegen in vierjährigem Abstand erhoben, wobei der Auswahlsatz zwischen 0,5 und 1 % variiert. Im Rahmen des Zusatzprogramms werden Angaben zur beruflichen Bildung, zur allgemeinen Weiterbildung, zu Migranten, zur Gesundheit, zu Behinderteneigenschaften sowie zur Altersvorsorge erfragt. Durch die Berücksichtigung aller Personen im Haushalt wird der Haushaltskontext erfasst und es besteht zudem die Möglichkeit, Personen und Haushalts- oder Familieninformationen zu verknüpfen, sodass Aspekte des sozialen Kontextes von individuellen Akteuren abgebildet werden können. Damit bietet der Mikrozensus mit seinem umfangreichen Merkmalskatalog ein reichhaltiges Potenzial für die Untersuchung sozial- und wirtschaftswissenschaftlicher Fragestellungen, das bereits seit einigen Jahren auch intensiv für wissenschaftliche Analysen genutzt wird (siehe hierzu Schimpl-Neimanns 2002).

Die spezifischen Stärken des Mikrozensus liegen neben dem großen Stichprobenumfang in der langfristigen Verfügbarkeit, die eine Analyse zeitlicher Entwicklungen von Anfang der siebziger Jahre bis zum aktuellen Rand erlaubt. Mit einem Auswahlsatz von 1 % der Gesamtbevölkerung ist der Stichprobenumfang um ein Vielfaches höher als bei sozialwissenschaftlichen Erhebungen und ermöglicht damit die Analyse von kleinen Subpopulationen wie z. B. Migranten, Selbstständigen, allein Erziehenden, nichtehelichen Lebensgemeinschaften oder Hochqualifizierten. Hinzu kommt, dass der Mikrozensus bedingt durch die gesetzliche Teilnahmepflicht nur in einem sehr geringen Umfang von Unit-Non-Response (2,5 bis 3 %) betroffen ist.

Darüber hinaus wird der Mikrozensus zurzeit in dem Kooperationsprojekt „Mikrozensus-Panel“⁴⁾ als Längsschnittdatensatz aufbereitet (siehe Zühlke 2001, Heidenreich 2002). Damit steht für die Analyse von personen- und haushaltsbezogenen Veränderungen erstmals eine Datenbasis zur Verfügung, die auf einer Erhebung mit Teilnahmepflicht basiert, eine sehr große Fallzahl umfasst und im Vergleich zu den freiwilligen Erhebungen der empirischen Sozialforschung eine sehr geringe Quote an Antwortausfällen aufweist. Ein zentrales Problem besteht jedoch darin, dass im Mikrozensus Haushalte und Personen, die zwischen zwei Erhebungsjahren fortziehen, nicht wieder befragt werden. Diese Ausfälle können dann zu Ergebnisverzerrungen führen, wenn die räumliche Mobilität mit denjenigen haushalts- und personenbezogenen Veränderungen zusammenhängt, die auf der Basis des Mikrozensus abgebildet werden sollen. Im Rahmen des oben genannten Kooperationsprojektes wurde dieses Problem detailliert analysiert, und es wurden Methoden entwickelt, um das Ausmaß der Verzerrungen möglichst gering zu halten (siehe Zühlke 2003, Basic/Marek/Rendtel 2005, Konold 2005, Konold in diesem Band). Das Ziel des Projektes besteht darin, die Daten des Mikrozensus-Panels im Verlauf des ersten Halbjahres 2006 für die Wissenschaft zugänglich zu machen.

Weitere Aspekte aus dem Bereich Familie, Bildung sowie Erwerbstätigkeit und Freizeit können auf der Grundlage der **Zeitbudgeterhebung** betrachtet werden, welche eine differenzierte Analyse des zeitlichen Aufwands für diese unterschiedlichen Lebensbereiche erlaubt. Zeitbudgeterhebungen wurden bislang zwei Mal auf freiwilliger Basis durchgeführt, und zwar in den Jahren 1991/1992 und 2001/2002. Befragt wurden im Rahmen dieser Erhebungen etwa 7 200 bzw.

4) An dem Projekt beteiligen sich neben dem Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen das Statistische Bundesamt, das Zentrum für Umfragen, Methoden und Analysen (ZUMA) und die Freie Universität Berlin.

1. Übersicht über die Mikrodaten zur Analyse des Arbeitsmarktes, des Bildungssystems sowie der Familiensituation				
Statistik/Bereich	Erhebungseinheiten	Vollerhebung/Stichprobe	Periodizität	Erhebungsschwerpunkte
Mikrozensus	Haushalte und Personen	Stichprobe (ca. 370.000 Haushalte und 820.000 Personen)	Jährlich	Erwerbstätigkeit und Bildung, Haushalts- und Familiensituation
Zeitbudgeterhebung	Private Haushalte und Personen	Stichprobe (ca. 7.200 Haushalte und 16.000 Personen (1991/92) sowie 5.400 Haushalte und 12.000 Personen (2001/02))	Unregelmäßig (Daten liegen für die Jahre 1991/92 und 2001/2002 vor)	Zeitaufwand für unterschiedliche Aktivitäten
Bevölkerungsstatistik	Eheschließungen Geburten	Vollerhebung	Jährlich	Eheschließungen: sozio-demographische Merkmale der Ehepartner Geburten: sozio-demographische Merkmale des Kindes und der Eltern
Europäische Erhebung zur beruflichen Weiterbildung	Unternehmen	Stichprobe (ca. 3.200 Unternehmen)	Unregelmäßig (Daten liegen für 1994 und 2000 vor)	Betriebliche Weiterbildungs-politik
Hochschulstatistik	Studierende Absolvent(inn)en Habilitierte Beschäftigte	Vollerhebung	Studierende und Absolvent(inn)en: Halbjährlich, Personal und Habilitationen: Jährlich	Studierende/Absolvent(inn)en: Sozio-demographische Merkmale, abgelegte Prüfungen Personal: sozio-demographische Merkmale sowie Angaben zur Vergütungsgruppe

5 400 Haushalte sowie die darin lebenden Personen. Aufgrund des breiten Kataloges an sozio-ökonomischen Merkmalen eignen sich die Zeitbudgeterhebungen zum einen für differenzierte Analysen der Zeitverwendung unterschiedlicher Bevölkerungsgruppen. Zum anderen können die Zusammenhänge zwischen unterschiedlichen Aktivitäten untersucht werden, wie z. B. der Zusammenhang zwischen familiärer Arbeitsteilung und Erwerbsbeteiligung. Die Daten beider Zeitbudgeterhebungen wurden bereits intensiv für wissenschaftliche Analysen genutzt (siehe Statistisches Bundesamt 2004).

Für die Analyse familiärer Prozesse können auch die Bevölkerungsstatistiken herangezogen werden. In der **Statistik der Eheschließungen** werden alle Ehen erfasst, die vor einem deutschen Standesbeamten geschlossen wurden. Neben dem Zeitpunkt der Eheschließung werden für beide Ehegatten eine Reihe weiterer Merkmale wie das Geburtsdatum, der bisherige Familienstand, das Vorhandensein gemeinsamer Kinder oder die Staatsangehörigkeit und die Religionszugehörigkeit erhoben. In der **Statistik der Geburten** sind alle Lebend- und Totgeburten erfasst, die sich in der Bundesrepublik im jeweiligen Berichtszeitraum ereignet haben und bei denen die Mutter zur Wohnbevölkerung Deutschlands zählt. Neben Informationen über das Geburtsdatum, das Geschlecht, die Staatsangehörigkeit, die Körperlänge oder das Geburtsgewicht des Kindes sind außerdem Angaben über Alter, Religionszugehörigkeit und die Staatsangehörigkeit der Eltern vorhanden. Für Geburten innerhalb bestehender Ehen sind darüber hinaus Informationen über vorherige Geburten verfügbar. Die Statistiken zu Eheschließungen und Geburten sind Vollerhebungen. Aufgrund der hohen Fallzahlen eignen sich beide Erhebungen sowohl für differenzierte Regionalanalysen als auch für die Untersuchung von Teilpopulationen. Im Gegensatz zu anderen Erhebungen enthalten die Bevölkerungsstatistiken allerdings keine Merkmale, mit denen die ökonomische Situation der Befragten beschrieben werden kann.

Für Analysen des Bildungssystems können im Mikrozensus vielfältige Informationen genutzt werden, die sich u. a. auf den allgemeinen und beruflichen Bildungsabschluss, den aktuellen Schulbesuch sowie die Teilnahme an allgemeinen und beruflichen Weiterbildungen beziehen. Darüber hinaus stehen mit der Europäischen Erhebung zur beruflichen Weiterbildung (CVTS) sowie der Hochschulstatistik Informationen zur Verfügung, die sich zur Analyse spezifischer Aspekte des Bildungsgeschehens eignen. Die **Europäische Erhebung zur beruflichen Weiterbildung** wurde in Deutschland auf freiwilliger Basis in den Jahren 1994 und 2000 durchgeführt (siehe Egner 2001). Für das Jahr 2000 liegen Angaben für knapp 3 200 deutsche Unternehmen vor. Die Erhebung umfasst Angaben zur Weiterbildungspolitik des Unternehmens und zu den verschiedenen durchgeführten Formen der Weiterbildung. Erfragt werden u. a. die Teilnahmezeiten an internen und externen Lehrveranstaltungen, die Themenschwerpunkte, Anbieter und Kosten dieser Lehrveranstaltungen sowie die Teilnahme an den verschiedenen Weiterbildungsformen differenziert nach Ungelernten/Angelernten, Fachkräften und Führungskräften.

Im Rahmen der **Hochschulstatistik** werden regelmäßig Informationen über die Studierenden, die abgelegten Prüfungen und Habilitationen, das Hochschulpersonal sowie die Finanzausstattung der Hochschulen gesammelt. Während die Angaben zu den Studierenden und Prüfungen zwei Mal jährlich erhoben werden, erfolgt die Erfassung der Habilitationen, des Hochschulpersonals und der Finanzausstattung ein Mal im Jahr. Im Rahmen der Studentenstatistik werden Angaben zur Person wie Geschlecht, Geburtsdatum, Staatsangehörigkeit, Semesterwohnsitz und

Heimatwohnsitz erfragt sowie Informationen zur Ersteinschreibung, zu früheren Auslandssemestern und zur Hochschulzugangsberechtigung gesammelt. Im Rahmen der Prüfungsstatistik werden für die Absolventen zusätzlich Angaben zur ersten bzw. zweiten Abschlussprüfung erhoben. Die Informationen aus der Studenten- und der Absolventenstatistik können miteinander kombiniert werden. Im Bereich der Habilitationsstatistik werden neben den Angaben zur Person das Fachgebiet sowie das Beschäftigungsverhältnis nachgewiesen. Die Personal- und Stellenstatistik erfasst die Angaben zur Person, zur organisatorischen Zugehörigkeit, zur fachlichen Zugehörigkeit sowie zum Beschäftigungsverhältnis. Da es sich bei den einzelnen Teilerhebungen der Hochschulstatistik um Vollerhebungen handelt, eignen sich diese Daten für differenzierte regionale und fachliche Analysen. Aufgrund der Aktualität der Daten können auch die Änderungen, die sich im Zuge der Hochschulreformen ergeben, untersucht werden.

3.2 Mikrodaten zur Analyse der Einkommenssituation privater Haushalte

Eine zentrale Datenquelle der amtlichen Statistik zur Analyse der Einkommenssituation privater Haushalte ist die vom Statistischen Bundesamt durchgeführte **Einkommens- und Verbrauchsstichprobe (EVS)**. Diese Erhebung wird auf freiwilliger Basis im Abstand von 5 Jahren bei denjenigen privaten Haushalten durchgeführt, deren monatliches Haushaltsnettoeinkommen weniger als 18 000 EUR beträgt.⁵⁾ Die Haushalte werden im Rahmen einer Quotenstichprobe gewonnen. Der angestrebte Stichprobenumfang betrug im Jahr 2003 etwa 74 000 Haushalte. Erhebungsschwerpunkte sind die Einkommen und Ausgaben, das Vermögen sowie die Schulden der Haushalte. Zudem wird die Ausstattung mit langlebigen Gebrauchsgütern erfragt. Ein Vorteil der Erhebung besteht zum einen im relativ großen Stichprobenumfang, der eine differenzierte Analyse der Einkommen und Ausgaben für unterschiedliche Haushaltstypen erlaubt. Zum anderen werden die Einnahmenquellen und die Verwendung der Ausgaben sehr differenziert erfasst, so dass auch hier detaillierte Analysen möglich sind. Die Daten der Einkommens- und Verbrauchsstichprobe wurden unter anderem im Rahmen des ersten und zweiten Armuts- und Reichtumsbericht der Bundesregierung intensiv analysiert (siehe Deutscher Bundestag 2001, 2005). Als Nachteil der Einkommens- und Verbrauchsstichprobe wird häufig angeführt, dass die Teilnahme an der Erhebung freiwillig ist. Es wird vermutet, dass die Einkommens- und Verbrauchsstichprobe – wie andere freiwillige Erhebungen auch – mit einem Mittelstandsbias verbunden ist (siehe z. B. Merz 2001). Aus diesem Grund werden für Einkommensanalysen häufig zwei weitere Statistiken herangezogen, und zwar die Einkommensteuerstatistik für die Analyse der hohen und die Sozialhilfestatistik für die Analyse der niedrigen Einkommen.

Die **Einkommensteuerstatistik** wird alle drei Jahre als Vollerhebung durchgeführt und enthält formal anonymisierte Angaben aus den Einkommensteuererklärungen aller Lohn- und Einkommensteuerpflichtigen.⁶⁾ Weiterhin werden die Lohnsteuerkarten derjenigen Lohnsteuerpflichtigen einbezogen, die weder zur Einkommensteuer-Veranlagung verpflichtet waren noch einen Antrag auf Veranlagung zur Einkommensteuer gestellt haben. Während somit sehr differenzierte Angaben aus dem Besteuerungsverfahren vorliegen, existieren nur wenige beschreibende

5) Die genannte Abschneidegrenze kam im Jahr 2003 zum Tragen. In früheren Erhebungsjahren lag sie etwas niedriger. – 6) Aufgrund der sehr hohen Fallzahlen wurden für statistikinterne Analysen Stichproben gezogen, die auch für wissenschaftliche Zwecke zur Verfügung gestellt werden. Siehe hierzu auch Zwick/Meyer 1998.

Merkmale, die eine Charakterisierung der Lohn- und Einkommensteuerpflichtigen erlauben. So sind beispielsweise im Bereich der sozio-demographischen Merkmale lediglich Informationen über das Geschlecht, die Religionszugehörigkeit und das Alter verfügbar. Weiterhin sind im Splittingfall einige Informationen über den Ehepartner bekannt. Mit Hilfe der Angaben zu den Kinderfreibeträgen können darüber hinaus Informationen über den Haushaltskontext rekonstruiert werden.

Um die niedrigen Einkommen abzubilden, wurden bisher aus dem Bereich der **Sozialhilfe-statistik** insbesondere die Angaben zu den Empfängern von laufender Hilfe zum Lebensunterhalt (HLU) genutzt.⁷⁾ Diese Statistik enthält alle Personen, die in dem betreffenden Berichtsjahr mindestens einen Monat laufende Hilfe zum Lebensunterhalt bezogen haben. Laufende Hilfe zum Lebensunterhalt wird denjenigen Personen gewährt, die aus eigener Kraft sowie durch Zahlungen anderer Sozialleistungsträger nicht in der Lage sind, ihren Lebensunterhalt zu bestreiten. Erfasst werden neben den Regelsätzen und gewährten Mehrbedarfen die tatsächlichen Kosten für Unterkunft und Heizung. Außerdem sind Informationen über das angerechnete Einkommen der Bedarfsgemeinschaft, der Nettoanspruch sowie der Zeitraum der Hilfestellung verfügbar. Weitere Merkmale beziehen sich auf sozio-ökonomische Angaben wie Alter und Geschlecht der Betroffenen sowie auf die Haushaltszusammensetzung. Die Daten über die Empfänger von laufender Hilfe zum Lebensunterhalt stehen bereits seit längerem in Form eines Public-Use-Files für wissenschaftliche Analysen zur Verfügung und werden insbesondere für Analysen im Bereich der Armutsberichterstattung genutzt. Im Forschungsdatenzentrum liegt das vollständige Datenmaterial vor, sodass das gesamte Analysepotenzial der Daten in regionaler und fachlicher Hinsicht ausgeschöpft werden kann.

Neben den bereits genannten Erhebungen wird in naher Zukunft die **Gemeinschaftsstatistik über Einkommen und Lebensbedingungen** für Einkommensanalysen zur Verfügung stehen. Bei dieser Erhebung handelt es sich um die Nachfolgeerhebung des Europäischen Haushaltspanels (siehe Körner/Meyer/Minkel/Timm 2005). Erstmals wurde die Gemeinschaftsstatistik über Einkommen und Lebensbedingungen in Deutschland unter der Bezeichnung „Leben in Europa“ im Jahr 2005 durchgeführt. Im Rahmen dieser Erhebung werden private Haushalte und Personen auf freiwilliger Basis befragt. Der angestrebte Stichprobenumfang beträgt 14 100 Haushalte. Erhebungsschwerpunkte sind die personen- und haushaltsbezogenen Einkommen, die Wohnsituation des Haushaltes, die Gesundheit der Befragten, Fragen zur Kinderbetreuung sowie die Einschätzung der eigenen finanziellen Lage. Zusätzlich gibt es jährlich wechselnde Erhebungsschwerpunkte. So werden im Jahr 2005 Fragen zur generationsübergreifenden Armut und im Jahr 2006 Fragen zur sozialen und kulturellen Teilhabe gestellt. Die Gemeinschaftsstatistik über Einkommen und Lebensbedingungen ist als Rotationspanel angelegt, sodass auf dieser Basis sowohl Quer- als auch Längsschnittinformationen gewonnen werden können.

⁷⁾ Der Empfängerkreis hat sich durch die im SGB II neu geregelte Grundsicherung für Arbeitssuchende, die zum 01. 01. 2005 in Kraft trat, stark geändert. Die Grundsicherung für Arbeitssuchende ersetzt die HLU für alle Bedarfsgemeinschaften mit mindestens einem erwerbsfähigen Leistungsberechtigten. Die Zahl der HLU-Empfänger wurde außerdem durch die am 01. 01. 2003 eingeführte bedarfsorientierte Grundsicherung beeinflusst, die eine eigenständige soziale Leistung zum Lebensunterhalt für Hilfebedürftige im Alter und bei Erwerbsminderung vorsieht.

2. Übersicht über die Mikrodaten zur Analyse der Einkommenssituation privater Haushalte					
Statistik/Bereich	Erhebungseinheiten	Vollerhebung/Stichprobe	Periodizität	Erhebungsschwerpunkte	
Einkommens- und Verbrauchsstichprobe	Private Haushalte und Personen mit einem monatlichen Haushaltsnettoeinkommen unter 18 000 EUR	Stichprobe (Stichprobensoll: ca. 74 000 Haushalte)	Fünfjährlich (zuletzt 2003)	Einkommen und Ausgaben, Vermögen und Schulden sowie Ausstattung mit langlebigen Gebrauchsgütern	
Lohn- und Einkommensteuerstatistik	Steuerpflichtige	10 %-Stichprobe auf Basis der Vollerhebung	Dreijährlich (zuletzt 2004)	Angaben aus dem Besteuerungsverfahren	
Sozialhilfestatistik	Personen, die mindestens einen Monat laufende Hilfe zum Lebensunterhalt (HLU) bezogen haben	Vollerhebung	Jährlich	Anspruch und Bruttobedarf, sonstige Einkommensquellen, sozio-ökonomische Merkmale der Leistungsempfänger, Informationen über den Haushaltskontext	
Gemeinschaftsstatistik über Einkommen und Lebensbedingungen	Private Haushalte und Personen	Stichprobe (angestrebte werden 14 100 Haushalte)	Jährlich	Einkommen, Wohnsituation, Gesundheit, Kinderbetreuung	

3.3 Mikrodaten zur Analyse der betrieblichen Situation

Für zentrale Kennziffern des betrieblichen Handelns wie Umsatz, Investitionen oder Anzahl der Beschäftigten werden in der amtlichen Statistik keine bereichsübergreifenden, sondern branchenspezifische Erhebungen durchgeführt. Zusätzlich werden für einige Wirtschaftszweige wie das Verarbeitende Gewerbe sowohl kurzfristige monatliche bzw. vierteljährliche Statistiken für Konjunkturprognosen als auch Strukturstatistiken in größeren zeitlichen Abständen erhoben. Im Ergebnis werden im Bereich der Wirtschaftsstatistiken sehr viele Einzelerhebungen mit dem Ziel durchgeführt, unterschiedliche Aspekte des wirtschaftlichen Handelns abzubilden.

Im Wirtschaftsbereich **Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden** (im Folgenden kurz „Verarbeitendes Gewerbe“) werden Betriebe befragt, die einem Unternehmen dieses Wirtschaftsbereiches angehören, das mehr als 20 Beschäftigte hat.⁸⁾ Ist ein Betrieb im Verarbeitenden Gewerbe tätig, gehört jedoch einem Unternehmen an, dessen wirtschaftlicher Schwerpunkt in einem anderen Bereich liegt, so wird dieser Betrieb dann befragt, wenn er mindestens 20 Personen beschäftigt. Die Betriebe berichten monatlich über die Anzahl der Beschäftigten, die geleisteten Arbeitsstunden, die Bruttolohn- und Gehaltsummen, die Umsätze sowie die Auftragseingänge (sog. Monatsbericht im Verarbeitenden Gewerbe). Vierteljährlich werden außerdem Informationen über die produzierten Waren (sog. Produktionserhebungen)⁹⁾ und jährlich über getätigte Investitionen (sog. Investitionserhebung) erhoben. Des Weiteren werden Mehrbetriebsunternehmen einmal jährlich nach den Unternehmensergebnissen befragt. Zudem wurde bis 2002 einmal jährlich eine Befragung bei Kleinbetrieben (1 bis 19 Beschäftigte) des Verarbeitenden Gewerbes durchgeführt, sodass eine vollständige Abbildung der betrieblichen Situation für diesen Wirtschaftszweig möglich war.¹⁰⁾ Die Daten des Monatsberichts im Verarbeitenden Gewerbe wurden bereits vor Einrichtung des Forschungsdatenzentrums im Rahmen des Projektes „Firmendaten aus der amtlichen Statistik (FiDAS)“ genutzt. Die Ergebnisse dieses Projektes sind in Pohl/Fischer/Rockmann/Semlinger 2003 dokumentiert.

In den Wirtschaftsbereichen **Handel¹¹⁾ und Gastgewerbe** werden – wie im Verarbeitenden Gewerbe – sowohl monatlich als auch jährlich Informationen erfasst. Bei den Monatserhebungen kommt dabei eine Abschneidegrenze zum Tragen, die sich am Umsatz des Unternehmens orientiert.¹²⁾ Monatlich werden der Umsatz sowie die Zahl der Voll- und Teilzeitbeschäftigten erhoben. Jährlich liegen Informationen über die Zahl der Arbeitsstätten, die tätigen Personen, den Personalaufwand, die Umsätze, die Vorleistungen, die Steuern, die Subventionen sowie die Investitionen vor. Im Einzelhandel wird darüber hinaus alle fünf Jahre die Zahl der Ladengeschäfte, die Verkaufsfläche sowie die Zahl der Marktstände erfragt. Im Gegensatz zu den monatlichen und jährlichen Erhebungen im Verarbeitenden Gewerbe wird in den Bereichen Handel und Gastgewerbe keine Vollerhebung, sondern lediglich eine Stichprobenerhebung durchgeführt.

8) Für ausgewählte Wirtschaftsbereiche ist die Erfassungsgrenze auf zehn und mehr tätige Personen herabgesetzt worden. Dies betrifft insbesondere Wirtschaftsbereiche, in denen kleinere Betriebe sehr häufig vertreten sind wie z. B. Fischverarbeitung, Kartoffelverarbeitung oder Herstellung von Futtermitteln. – 9) Die Erhebung erfolgt lediglich bei einer Stichprobe. Bei größeren Unternehmen werden auch diese Informationen monatlich erhoben. – 10) Zukünftig sollen diese Informationen aus dem Unternehmensregister gewonnen werden. – 11) Unter Handel werden im Folgenden die Bereiche „Handel, Instandhaltung und Reparatur von Kraftfahrzeugen und Gebrauchsgütern“, „Handelsvermittlung und Großhandel“ und „Einzelhandel und Reparatur von Gebrauchsgütern“ verstanden. – 12) Es wird der Jahresumsatz ohne Umsatzsteuer zu Grunde gelegt. In den einzelnen Bereichen kommen folgende Abschneidegrenzen zum Tragen: Einzel- und Kraftfahrzeughandel (einschließlich Instandhaltung, Reparatur) 250 000 EUR, Großhandel 1 000 000 EUR, Handelsvermittlung und Gastgewerbe 50 000 EUR.

3. Übersicht über die Mikrodaten zur Analyse der betrieblichen Situation				
Statistik/Bereich	Erhebungseinheiten	Vollerhebung/Stichprobe	Periodizität	Erhebungsschwerpunkte
Erhebungen im Verarbeitenden Gewerbe	Betriebe (und Unternehmen)	Vollerhebung mit Abschneidegrenze	Monatlich	Beschäftigte, geleistete Arbeitsstunden, Brutto- und Gehaltssummen, Umsätze und Auftragseingänge
		Stichprobenerhebung mit Abschneidegrenze (höchstens 43 000 Betriebe)	Vierteljährlich (bei größeren Betrieben monatlich)	Produzierte Waren
		Vollerhebung mit Abschneidegrenze	Jährlich	Investitionen, ggf. Unternehmensergebnisse
Erhebungen im Handel und Gastgewerbe	Unternehmen	Stichprobenerhebung mit Abschneidegrenze (höchstens 40 000 (Handel) bzw. 10 000 (Gastgewerbe) Unternehmen)	Monatlich	Umsatz, Zahl der Vollzeit- und Teilzeitbeschäftigten
		Stichprobenerhebung (höchstens 55 000 (Handel) bzw. 12 000 (Gastgewerbe) Unternehmen)	Jährlich	Zahl der Arbeitsstätten, Tätige Personen, Personalaufwand, Umsätze, Vorleistungen, Steuern, Subventionen, Investitionen
		Stichprobenerhebung (höchstens 55 000 Unternehmen (Handel))	Fünfjährlich (nur im Handel)	Zahl der Ladengeschäfte, Verkaufsfläche, Marktstände
Strukturerhebung im Dienstleistungsbereich	Unternehmen und Einrichtungen zur Ausübung freiberuflicher Tätigkeiten	Stichprobenerhebung bei etwa 15 % aller Auskunftsspflichtigen (für das Berichtsjahr 2000 ca. 75 000 Unternehmen)	Jährlich	Tätige Personen, Löhne und Gehälter, Umsätze, Vorleistungen, Steuern, Subventionen und Investitionen
		Stichprobenerhebung bei etwa 10 % der Auskunftsspflichtigen mit Abschneidegrenze (2002: ca. 22 000 Betriebe; ca. 845 000 Beschäftigte)	Unregelmäßig (Daten liegen für 2002 vor)	Bruttomonatsverdienste, gesetzliche Abzüge, Zulagen, individuelle Merkmale der Arbeitnehmer/-innen, Merkmale des Arbeitsplatzes sowie des Betriebes

Die Erhebungen im Handel und im Gastgewerbe waren bis zum Berichtsjahr 2000 die einzigen amtlichen Primärerhebungen bei Unternehmen, auf deren Basis Aussagen über den Dienstleistungsbereich gemacht werden konnten. Der Dienstleistungsbereich war also nur unvollständig repräsentiert. Geändert hat sich diese Situation mit der Einführung der **Strukturerhebung im Dienstleistungsbereich**, die erstmals für das Berichtsjahr 2000 durchgeführt wurde (siehe Petruschke 2002). Neben den Informationen über den Handel und das Gastgewerbe liegen damit auch Informationen für die Bereiche „Verkehr und Nachrichtenübermittlung“ sowie „Grundstücks- und Wohnungswesen, Vermietung beweglicher Sachen und Erbringung von Dienstleistungen überwiegend für Unternehmen“ vor. Berücksichtigt werden in dieser Statistik Unternehmen und Einrichtungen, die der Umsatzsteuerpflicht unterliegen. Die Dienstleistungsstatistik wird einmal jährlich auf der Basis einer Stichprobenerhebung durchgeführt, zu der höchstens 15 % aller Auskunftspflichtigen herangezogen werden dürfen. Dabei werden die Anzahl der tätigen Personen, die Löhne und Gehälter, die Umsätze, die Vorleistungen, die Steuern sowie die Subventionen und Investitionen erfragt.

Die genannten Erhebungen zielen insbesondere auf die zentralen Indikatoren Umsatz, Beschäftigte und Investitionen ab. Daneben werden in der amtlichen Statistik weitere bereichsübergreifende Erhebungen im Verarbeitenden Gewerbe und im Dienstleistungssektor durchgeführt, die sich mit der Struktur der Löhne und Gehälter in den Unternehmen befassen. Eine dieser Erhebungen ist die **Gehalts- und Lohnstrukturerhebung**, die unregelmäßig bei etwa 10 % der Auskunftspflichtigen durchgeführt wird. Für sie werden Angaben über die Bruttomonatsverdienste, die gesetzlichen Abzüge, Zulagen, individuelle Merkmale der Arbeitnehmer/-innen, Merkmale des Arbeitsplatzes sowie des Betriebes erhoben. Für wissenschaftliche Analysen ist die Gehalts- und Lohnstrukturerhebung von besonderem Interesse, da sie als so genannter „Employer-Employeedatensatz“ konzipiert ist. Dies bedeutet, dass in dem Datensatz sowohl Angaben zu den einzelnen Beschäftigten als auch zu den Betrieben vorhanden sind. Diese Datenbasis kann daher genutzt werden, um zum Beispiel Unterschiede zwischen Löhnen und Gehältern in Abhängigkeit von dem betrieblichen Kontext und der dort vorhandenen Arbeitsplatzstruktur zu analysieren.¹³⁾

4 Zusammenfassung und Ausblick

Im Rahmen des vorliegenden Beitrages wurden die Aufgaben des Forschungsdatenzentrums der Statistischen Landesämter vorgestellt sowie die Zugangswege und das Datenangebot beschrieben. In den vergangenen zwei Jahren konnte eine Vielzahl an Mikrodatenbeständen für die wissenschaftliche Forschung zur Verfügung gestellt werden, wodurch sich die Dateninfrastruktur deutlich verbessert hat. Das Forschungsdatenzentrum der Statistischen Landesämter wird das Datenangebot sukzessive durch die Integration weiterer Erhebungsjahre und Statistiken ausbauen, wobei hierfür die Nutzerinteressen von entscheidender Bedeutung sind. Für die Wissenschaft bestehen unterschiedliche Möglichkeiten des Datenzugangs, die bereits intensiv genutzt werden. Insgesamt wurden in den Jahren 2004 und 2005 für über 80 Forschungsprojekte Daten beim Forschungsdatenzentrum der Statistischen Landesämter beantragt. Hinzu kommen etwa 80 Forschungsprojekte, für die die Daten der statistischen Landesämter durch das Statistische Bundesamt bereitgestellt worden sind. Es ist zu erwarten, dass mit dem weiteren

13) Zur Gehalts- und Lohnstrukturerhebung vgl. auch den Beitrag von Hafner in diesem Band.

Aufbau des Datenangebotes die Nachfrage nach amtlichen Mikrodaten noch weiter steigen wird. Im Ergebnis wird das Datenangebot der amtlichen Statistik durch die Einrichtung von Forschungsdatenzentren für wissenschaftliche Analysen wesentlich intensiver genutzt, als dies zuvor der Fall war. Somit können Erkenntnisse erzielt werden, die ansonsten in dieser Form nicht hätten gewonnen werden können.

Literatur

Basic, Edin/Marek, Ivo/Rendtel, Ulrich (2005): The German Microcensus as a Tool for Longitudinal Data Analysis: An Evaluation Using SOEP Data; in: Schmollers Jahrbuch, Heft 1, 1 – 16.

Egner, Ute (2001): Zweite Europäische Erhebung zur beruflichen Weiterbildung (CVTS2): Methodik und erste Ergebnisse; in: Wirtschaft und Statistik Heft 12, 1008 – 1022.

Deutscher Bundestag (2001): Lebenslagen in Deutschland – Erster Armuts- und Reichtumsbericht. Drucksache 14/5990 vom 8. Mai 2001, Berlin 2001.

Deutscher Bundestag (2005): Lebenslagen in Deutschland – Zweiter Armuts- und Reichtumsbericht. Drucksache 15/5015 vom 3. März 2005, Berlin 2005.

Forschungsdatenzentrum der Statistischen Landesämter (2006): Amtliche Mikrodaten für die Agrar- und Umweltwissenschaften – Beiträge zu den Nutzerkonferenzen des FDZ der Statistischen Landesämter 2005

Hafner, Hans-Peter (2006): Erhebungen zu Verdiensten und Arbeitskosten. In diesem Band.

Heidenreich, Hans-Joachim (2002): Längsschnittdaten aus dem Mikrozensus: Basis für neue Analysemöglichkeiten; in: Allgemeines Statistisches Archiv, Band 86, Heft 2, 213 – 231.

Körner, Thomas/Meyer, Iris/Minkel, Hartmut/Timm, Ulrike (2005): Leben in Europa – Die neue Statistik über Einkommen und Lebensbedingungen; in: Wirtschaft und Statistik Heft 11, 1137 – 1152.

Konold, Michael (2005): Der Ausfall räumlich mobiler Personen im Mikrozensus-Panel und seine Auswirkungen auf Übergangsanalysen im Familien- und Erwerbsbereich. Ergebnisse empirischer Untersuchungen. Unveröffentlichtes Arbeitspapier. Düsseldorf.

Konold Michael (2006): Möglichkeiten der Analyse von Arbeitsmarktübergängen mit Daten des Mikrozensus-Panels; in diesem Band.

Merz, Joachim (2001): Was fehlt in der EVS? Eine Verteilungsanalyse hoher Einkommen mit der verknüpften Einkommensteuerstatistik für Selbständige und abhängig Beschäftigte. Diskussionspapier Nr. 30, Universität Lüneburg, Lüneburg.

Petrauschke, Bernd (2002): Erste Unternehmenserhebung im Dienstleistungsbereich; in: Wirtschaft und Statistik Heft 11, 918 – 927.

Philipp, Gabriele/Schoffer, Olaf (2006): Krankenhaus-Diagnosestatistik – Scientific-Use-File und beispielhafte Analysen; in diesem Band.

Pohl, Ramona/Fischer, J./Rockmann, Ulrike/Semlinger, Klaus (2003): Analysen zur regionalen Industrieentwicklung – Sonderauswertungen einzelbetrieblicher Daten der amtlichen Statistik. FHTW Berlin und Statistisches Landesamt Berlin.

Ronning, Gerd/Gnoss, Roland (2003): Anonymisierung wirtschaftsstatistischer Einzeldaten. Beiträge zum Workshop am 20./21. März 2003 in Tübingen. Forum der Bundesstatistik Band 42, Statistisches Bundesamt, Wiesbaden.

Schimpl-Neimanns, Bernhard (2002): Anwendungen und Erfahrungen mit dem Scientific-Use-File des Mikrozensus. ZUMA-Arbeitsbericht 2002/01.

Statistisches Bundesamt (Hrsg.) (2004): Alltag in Deutschland: Analysen zur Zeitverwendung. Beiträge zur Ergebniskonferenz der Zeitbudgeterhebung 2001/02 am 16./17. Februar 2004 in Wiesbaden. Forum der Bundesstatistik Band 43, Wiesbaden.

Sturm, Roland/Lenz, Rainer (2006): Erste Scientific-Use-Files aus den Wirtschaftsstatistiken; in diesem Band.

Wagner, Gerd G./Wagner, Joachim (2005): Workshop „Untersuchungen mit Mikrodaten aus der Amtlichen Wirtschafts- und Sozialstatistik“; in: Schmollers Jahrbuch, Journal of Applied Social Science Studies, Zeitschrift für Wirtschafts- und Sozialwissenschaften, 125. Jahrgang 2005 Heft 4.

Zühlke, Sylvia (2001): Längsschnittanalysen auf der Basis des Mikrozensus. Methodische Probleme und Lösungsansätze. In: Statistische Analysen und Studien Nordrhein-Westfalen, Heft 4, 3 – 13.

Zühlke, Sylvia (2003): Systematische Ausfälle im Mikrozensus-Panel: Ausmaß und Auswirkungen auf die Qualität von Arbeitsmarktanalysen; in: Allgemeines Statistisches Archiv, Band 87, Heft 1, 39 – 58.

Zühlke, Sylvia/Hetke, Uwe (2003): Datenbedarf und Datenzugang: Ergebnisse der ersten Nutzerbefragung des Forschungsdatenzentrums der Statistischen Landesämter; in: Allgemeines Statistisches Archiv, Band 87, 321 – 334.

Zühlke, Sylvia/Zwick, Markus/Scharnhorst, Sebastian/Wende, Thomas (2005): Die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder; in: ZA-Informationen 56, 168 – 182.

Zwick, Markus/Meyer, Kurt (1998): Einzeldatenmaterial und Stichproben innerhalb der Steuerstatistiken; in: Wirtschaft und Statistik Heft 7, 566 – 573.

Autoren- und Kontaktangaben:

Dr. Sylvia Zühlke, FDZ der Statistischen Landesämter, Landesamt für Datenverarbeitung und Statistik NRW, E-Mail: Sylvia.Zuehlke@lds.nrw.de

Helga Christians, FDZ der Statistischen Landesämter, Landesamt für Datenverarbeitung und Statistik NRW, E-Mail: Helga.Christians@lds.nrw.de

The Effects of Public Schooling Expenditures on Individual Labour Market Outcomes*)

Andreas Ammermüller, Anja Kuckulenz, Thomas Zwick

Our empirical analyses show the effects of public schooling expenditures on individual labour market outcomes in the German states and highlight differences between East and West Germany. When we consider East and West Germany separately, we find a positive and significant correlation between GDP per capita and schooling attainment. There is an increase in educational attainment in (West) Germany induced by younger cohorts. Educational attainment of younger cohorts in East Germany has fallen below the national average in recent years.

Schooling has an important influence on individual wages. An additional year of schooling results on average in an increase of wages by 7.8 percent. The returns to schooling are lower in East than in West Germany. The marginal employment effects of schooling, however, are higher in East than in West German states, where the qualification background seems to be less important for finding a job. It is therefore essential to consider not only the wage but also the employment effect of schooling if the impact of education on the labour market outcomes under different labour market circumstances is analysed.

1 Introduction

The role and importance of public education spending has been discussed by policymakers as means to enhance both individual and collective productivity, and consequently to secure economic performance and social prospects in the long run. Contributing to this discussion, we provide empirical evidence on the relation between education expenditures and attainment in the German states. In addition, we present some econometric estimates on the effects of schooling on individual wages and employment (compare de la Fuente et al., 2003 for Spain).

In view of the financing structure of education in Germany,¹⁾ most variation in public expenditure for education can be expected between the 16 states. We indeed find large differences in expenditures for education between the German states. Most pupils attend public schools that are free of charge and also all institutions of higher education are basically free.²⁾ Therefore, only public expenditures are taken into account.

*) We are grateful to Angel de la Fuente and Vittorio Campanelli for useful comments and suggestions and Michael Gebel, Franziska Mientus, and Felix Pannenberg for research assistance. – 1) Public financing of education is split between the federal, state and municipal level in Germany. For schools, about 80 percent of the financial means are provided by the states, while municipalities account for the remaining 20 percent, mainly for maintaining school buildings. The states are responsible for the financing of universities, except for additional grants from the federal level for new buildings and large purchases. – 2) In Germany, the share of private education expenditures is higher than in comparable countries – while 4.5 % of GDP is invested by the state, private entities invest 1.2 % of GDP. The private investments are mainly concentrated on the dual apprenticeship system, the pre-school education of children and continuing training on the workplace (Klöß and Weiß, 2003). The share of students on private schools (8 %) and universities (2 %) is tiny (Statistisches Bundesamt). In addition, even private schools and universities frequently receive considerable public funding.

We show that there is a correlation between relative regional financing per head and relative income per capita separately for secondary and tertiary schooling. Then we show econometric estimates of the effects of schooling on wages. The estimated Mincerian returns in the different German states are then related to relative educational attainment in these states, to the states' relative gross domestic product (GDP) per capita and to labour market participation and employment. In order to give a full picture of labour market consequences of education, we add the effects of schooling on employment chances. This has been rarely done in the literature on the German labour market. So far, mainly the wage effect of schooling has been analysed, while the impact on employment is often ignored (Wolter and Weber, 1999; Zwick, 2001).

2 Educational expenditure and attainment in the German states

Regional expenditure on secondary and university education

Average public expenditure per student per year is divided into expenditure in secondary schooling and university.³⁾ Expenditures are total expenditures per student net of investments and weighted averages with weights of 2/3 for secondary and 1/3 for university. This follows the assumption that half of secondary schooling graduates will go on to university.

1. Average public expenditure per student, 2000			
	Secondary	University	Combined
Saxony-Anhalt	104.3	157.2	135.7
Saarland	95.7	155.6	131.2
Schleswig-Holstein	100.0	148.4	128.7
Mecklenburg-W. P.	91.3	148.0	124.9
Baden-Wuerttemberg	100.0	129.5	117.5
Hamburg	143.5	98.5	116.8
Bavaria	106.5	120.9	115.0
Berlin	115.2	112.3	113.5
Thuringia	106.5	112.5	110.1
Saxony	91.3	114.0	104.8
Lower Saxony	100.0	97.6	98.6
Hesse	97.8	90.8	93.6
Rhineland-Palatinate	97.8	80.4	87.5
Bremen	119.6	59.2	83.7
North Rhine-Westphalia	95.7	72.2	81.7
Brandenburg	89.1	57.2	70.2
Average	100	100	100
Average euros	4600.0	13628.5	7579.4
Average East	98.2	115.2	103.8
Average West	100.4	96.3	99.0

Note: Expenditure per student in secondary schooling is taken from „Bildung im Zahlenspiegel“ 2003; expenditure per student in university is calculated by dividing the total current expenditures of universities by the number of students inscribed at university (source: Statistisches Bundesamt).

In Table 1, the normalized average expenditures per student in 2000 are presented, with German average expenditure per student equal to 100. The states are sorted by combined expenditure per student and East German states are in bold type. For expenditure in secondary schooling, the variation between states is rather low. The city states of Hamburg, Bremen and Berlin have by far the highest expenditures per students in secondary schooling. This is mostly due to higher overhead costs per student and does not imply that students directly receive more resources (Ammermüller and Weber, 2005). Expenditures in secondary schooling are slightly higher

3) Secondary schooling includes all school types except vocational education. University includes all tertiary educational institutions such as higher technical colleges (Fachhochschule).

in Western than in Eastern states (Berlin, Saxony-Anhalt, Mecklenburg-Western Pommerania, Thuringia, Saxony, and Brandenburg).

At the university level, expenditures per student are almost three times as high as in secondary schooling. The variation in expenditures is also much higher. Eastern states have on average a higher expenditure per student than Western states. It is well-known that universities in East Germany are often well-equipped but less popular among students. They also have a higher ratio of university employees to student⁴). Due to the higher average expenditure in university compared to secondary schooling, Eastern states also have on average higher combined expenditures per student.

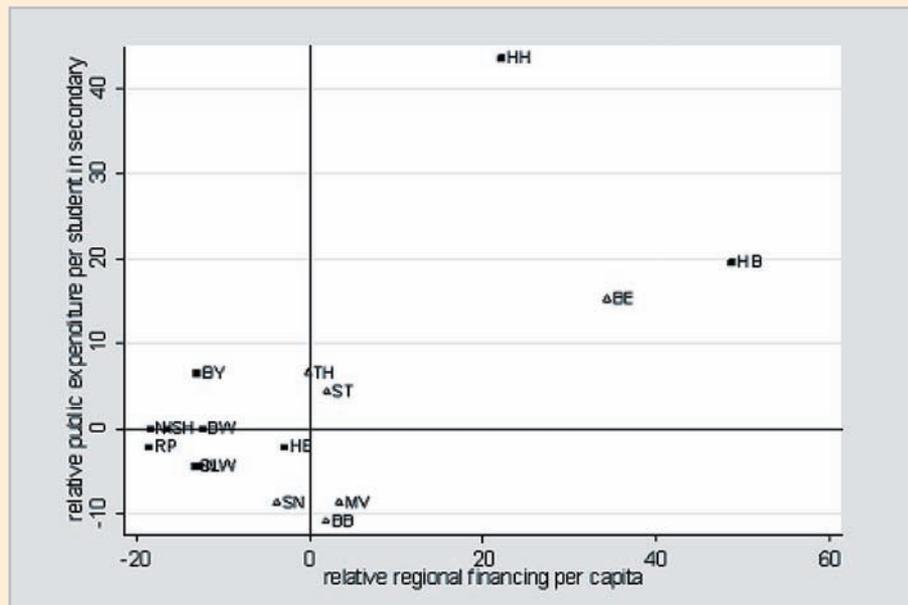
While combined expenditure per student is higher in many Eastern states, it is neither correlated with regional financing per capita nor with GDP per capita. This is shown in the regression below with t-values in brackets:

$$1) \text{ combined expenditures/student} = -0.098 * \text{regional financing/capital} - 0.038 * \text{GDP/capital} + c$$

(t-values) (0.34) (0.20)

with an R² of 0.017⁵). When combined expenditure per student is replaced by expenditure in secondary schooling as the dependent variable, both resource measures are highly significantly correlated with expenditure per student in secondary schooling, however. This is shown in the following regression:

Fig. 1 Public expenditure per pupil in school vs. regional financing per capita in 2000



Key: BW = Baden-Wuerttemberg, BY = Bavaria, BE = Berlin, BB = Brandenburg, HB = Bremen, HH = Hamburg, HE = Hesse, NI = Lower Saxony, MV = Mecklenburg-Western Pommerania, NW = North Rhine-Westphalia, RP = Rhineland-Palatinate, SL = Saarland, SN = Saxony, ST = Saxony-Anhalt, SH = Schleswig-Holstein, TH = Thuringia. Eastern states are denoted by a triangle, Western states by a square.

4) Statistisches Bundesamt, data for 2000. – 5) All variables are measured in percentage deviations from the German average. Including a dummy for Eastern states or estimating equation (1) separately for Eastern and Western states does not lead to significantly positive coefficients.

Our data base for measuring educational attainment on the state level is the yearly micro-census. The data set comprises 1 % of the population living in Germany and is representative for Germany. We know the highest educational attainment of the individual and can therefore calculate the individual years of schooling as well as the fraction of the population that has completed different levels of education.

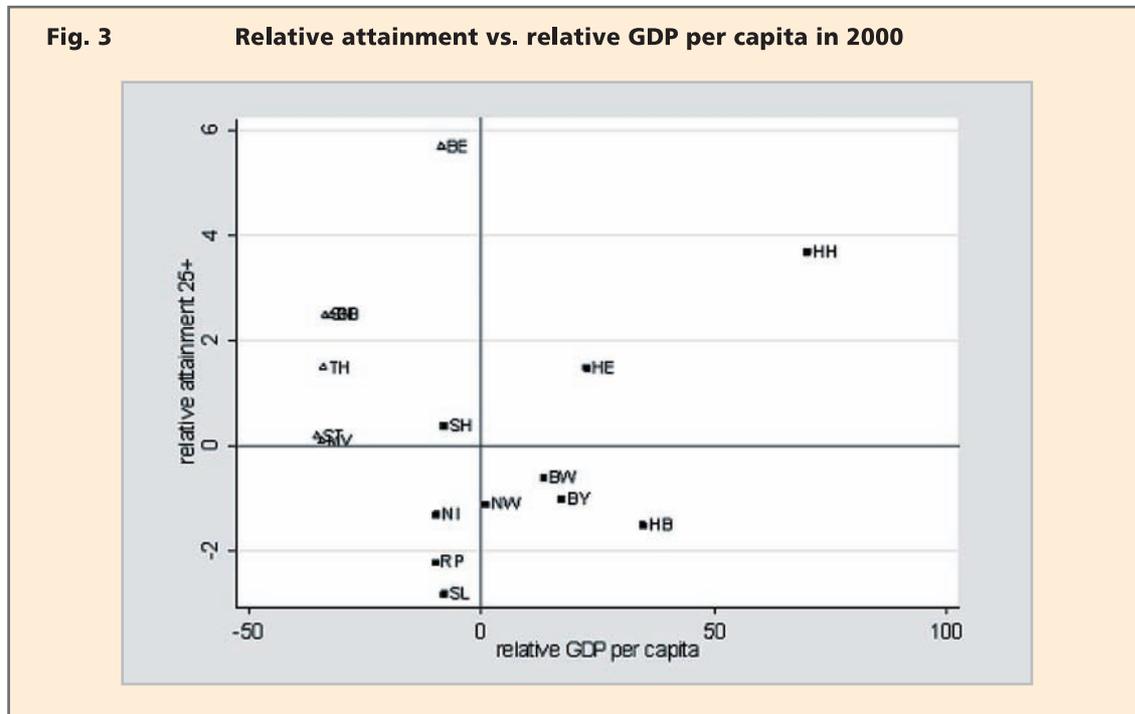


Figure 3 plots relative educational attainment of the population older than 25 years of age measured in years of schooling and relative GDP per capita for the year 2000 per state. No positive relationship between the two variables can be detected in the scatter plot. When we include a dummy for the East German states that have a relatively high educational attainment but low GDP per capita in a regression of relative attainment on GDP per capita, the relationship turns out to be significantly positive at the one-percent-level, however:

$$3) \text{ relative attainment} = 0.066 * \text{GDP / capital} + 5.345 * \text{East} + c$$

(t-values) (3.73) (5.12)

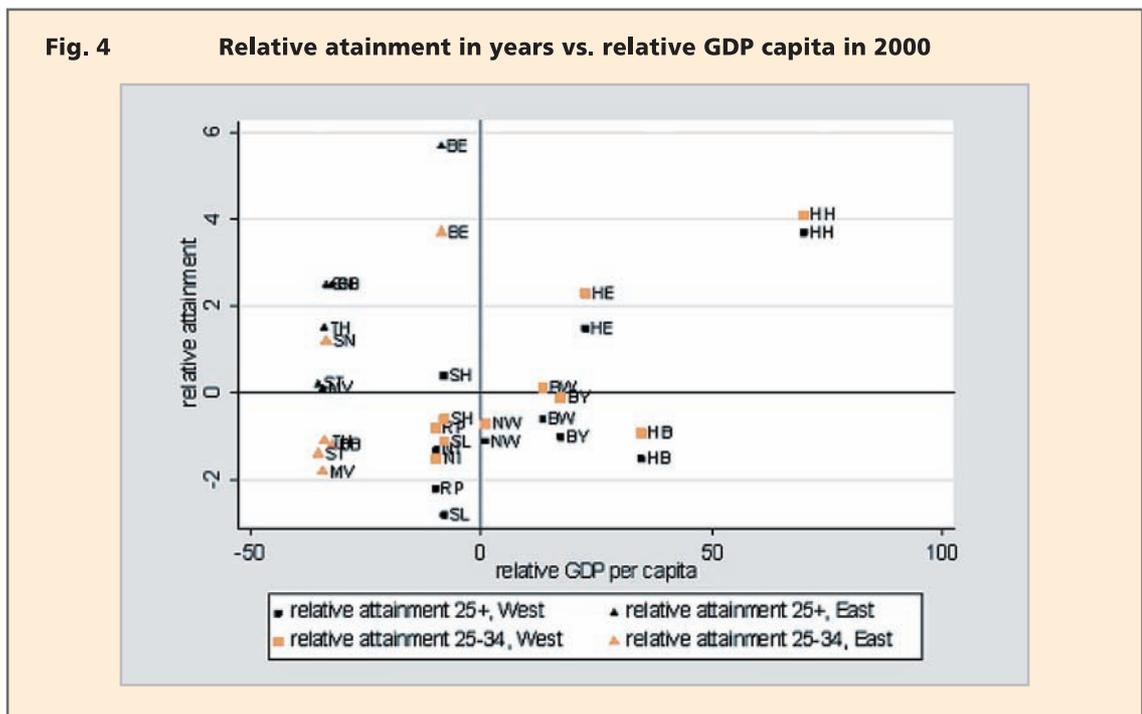
Therefore states with above-average GDP have above-average educational attainment.

We next analyse the developments in educational convergence across generations in the various German regions in order to derive predictions from the observed trend with regard to the future evolution of regional disparities. Table 2 uses average years of schooling completed as an indicator of educational attainment. For each cohort, the average years of schooling have been normalised to 100. The first column refers to the whole population above 25 years of age. The values, above or below 100, reported for each region therefore indicate how the region considered is situated compared to the German average in the corresponding cohort. Moreover, at the bottom of the table, the average years of schooling completed in Germany are reported for each age cohort as well as the degree of regional dispersion in the cohort, measured by the coefficient of variation of relative educational attainment in the regions.

2. Average years of schooling by cohort, 2000 (Germany = 100)						
	25+	25 – 34	35 – 44	45 – 54	55 – 64	65+
Berlin	105.7	103.7	104.5	105.6	106.0	107.9
Hamburg	103.7	104.1	103.4	102.4	102.6	103.9
Brandenburg	102.5	98.8	101.4	104.0	104.7	102.7
Saxony	102.5	101.2	101.6	103.8	105.3	102.8
Hesse	101.5	102.3	101.7	100.8	100.9	100.9
Thuringia	101.5	98.9	101.1	102.2	103.9	101.9
Schleswig-Holstein	100.4	99.4	99.8	100.7	101.5	101.0
Saxony-Anhalt	100.2	98.6	99.3	101.6	103.0	100.8
Mecklenburg-West. Pom.	100.1	98.2	98.1	101.8	102.9	100.2
Baden-Wuerttemberg	99.4	100.1	99.8	99.1	98.0	99.2
Bavaria	99.0	99.9	99.4	98.2	98.0	97.9
North Rhine-Westphalia	98.9	99.3	99.2	98.6	98.2	99.7
Lower Saxony	98.7	98.5	99.3	99.0	98.7	98.3
Bremen	98.5	99.1	98.0	97.6	99.5	99.9
Rhineland-Palatinate	97.8	99.2	98.3	97.9	97.1	96.8
Saarland	97.2	98.9	96.2	97.8	96.7	98.1
Germany	100.0	100.0	100.0	100.0	100.0	100.0
Germany (years)	12.1	12.7	12.6	12.4	11.9	11.1
coeff. of var.	2.2 %	1.8 %	2.1 %	2.4 %	3.0 %	2.6 %

Source: Micro-census 2000, own computations.

It can be seen that educational attainment increased continuously and significantly from the older to the younger generations, though not dramatically. The regional disparities in educational attainment are rather low – the difference in relative attainment is less than 10 points between top and bottom regions, and they have slightly decreased across generations. Therefore, through the renewal of generations alone, and assuming that future generations are best comparable with the youngest cohort, average educational attainment in the population will further increase and regional disparities will slowly decrease in the future. This phenomenon is expected to be more pronounced, however, if the trends towards increasing education and decreasing regional dispersion continue for future (younger) cohorts.



Regarding educational attainment, the ranking of the regions has not changed substantially between older and younger generations. The only noticeable change is that educational attainment in Eastern states used to be above German average for older generations and has fallen below German average for younger generations. Figure 4 plots relative educational attainment (percentage deviation of the regional educational attainment from the German average in the corresponding age group) of the entire adult population (above 25) and of the youngest cohort (aged 25 to 34) against relative GDP per capita of the region (percentage deviation of regional GDP from the German average GDP). A striking feature is that there is no obvious correlation between relative income and relative educational attainment. However, a clear pattern emerges: richer regions (GDP per capita above average) proved successful in improving the relative educational attainment for the youngest generations whereas in poorer regions (mainly Eastern states), the relative educational attainment of the youngest generation dropped. Since education is an essential component of human capital and therefore also an essential factor of economic development, this is a bad sign for future economic convergence across regions, in particular between (richer) West and (poorer) East Germany.

3 Econometric estimates of the effects of schooling

In this section, we present estimates of the Mincerian returns to schooling from wage regressions including years of schooling as explanatory variables. These estimates for the base year 2000 are related to relative educational attainment and to relative GDP per capita of the German states in order to see, whether qualification supply on the labour market and economic welfare have an impact on individual returns to schooling.

Schooling and wages

To estimate the yearly returns to schooling for individual j aggregated to German states level, we use the micro-census. We apply the standard Mincerian wage equation including the years of schooling (S), potential experience ($potexp$), potential experience squared ($potexp^2$) and a dummy for sex ($Dsex$). The dependent variable is log of net hourly wages $\ln w$.

$$4) \ln w_i = c + \theta S_i + a*potexp_i + b*potexp_i^2 + d*Dsex_i + u_i.$$

In the micro-census we have information about the highest educational level obtained. This allows us to calculate the cumulative duration, i. e. the years of schooling. Since we do not have information about the actual years of schooling, we use the official durations of each educational level (see Table 3).

3. Attainment level used in the micro-census and cumulative durations	
	Cumulative duration
No degree	8
Lower secondary degree (low.sec.) – Hauptschule	9
Middle secondary degree (mid.sec.) – Realschule	10
Lower sec. degree + apprenticeship (l.sec. + app.)	12
Middle sec. degree + apprenticeship (m.sec. + app.)	12
Secondary and vocational school (sec. + voc.)	12
Higher secondary degree (h.sec.) – Abitur	13
Higher sec. degree + apprenticeship (h.sec. + app.)	15
Technical College (tech. collage)	17
University degree (univ.)	18

The estimated yearly effects of schooling estimated by the wage equation (4) for each state for the year 2000 are presented in Table 4. The estimated values of parameter θ , measuring the effect of an additional year of schooling on wages, are depicted in the first column. The other columns present the estimated yearly effects of an additional year of schooling for a specific educational level. Instead of using years of schooling, dummies for educational levels have been introduced into the wage equation (4). The yearly effects are calculated as the difference in coefficients between two adjacent educational degrees divided by the additional years that are needed to complete the educational cycle on average.

4. Estimated yearly effects of schooling in wage equations for the year 2000										
	θ	low. sec.	mid. sec.	l. sec. + app.	m. sec. + app.	sec. + voc.	h. sec.	h. sec. + app.	tech. college	univ.
Saarland	9.57	8.54	48.73	5.04	0.00	3.21	0.00	0.00	7.68	12.56
Rhinel.-Palatinate	9.18	17.97	36.34	7.28	0.00	5.20	7.53	0.00	5.06	6.11
Bavaria	8.71	18.30	26.86	5.11	4.50	9.18	4.44	5.90	8.30	7.93
Hesse	8.70	0.00	16.01	3.55	7.50	12.55	8.79	3.58	7.43	6.58
Schleswig-Holst.	8.17	0.00	14.46	1.73	1.37	6.71	5.51	6.25	8.25	6.95
Brandenburg	8.01	0.00	22.34	1.96	0.00	2.89	10.72	0.00	0.20	3.71
Baden-Wuerttem.	7.89	19.30	21.01	4.94	4.40	9.47	4.21	3.94	8.52	7.34
Bremen	7.83	17.44	16.92	4.62	4.69	5.97	0.63	21.67	9.95	11.07
Hamburg	7.69	7.22	26.51	5.01	1.91	1.35	6.05	3.99	4.73	5.89
North Rh.-Westp.	7.62	15.44	14.05	6.07	6.48	13.12	7.05	1.86	7.04	7.07
Saxony	7.16	10.08	0.00	0.00	0.00	11.16	9.66	0.00	1.24	4.75
Lower Saxony	7.12	30.95	15.13	3.38	3.96	9.96	3.10	8.00	7.22	8.30
Saxony-Anhalt	6.69	29.14	19.33	0.32	0.00	2.03	9.14	0.00	0.00	4.39
Berlin	6.62	0.00	0.24	3.78	6.74	12.32	8.80	4.44	4.78	6.10
Thuringia	6.34	122.38	0.00	0.00	0.00	4.19	3.69	0.00	4.60	6.27
Meckl.-West. P.	6.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.68	4.66
Germany	7.80	13.32	20.57	5.17	3.77	10.31	6.03	2.61	6.67	7.00

Note: Estimates from wage equation (4), using data from micro-census 2000.

The estimated average returns to an additional year of schooling in Germany are 7.8 percentage points, which can be seen as a weighted average of all estimated parameters of the German states. The Mincerian returns to schooling range between 6.15 percentage points in Mecklenburg-Western Pommern and 9.57 percentage points in Saarland. These estimates are comparable to previous estimates for Germany (Ammermüller and Weber, 2005; Lauer und Steiner, 2000). For educational levels, the yearly returns range between zero and 122 percent. However, these returns have to be interpreted relative to the prior educational level. Moreover, for some levels in smaller states, the number of observations was quite low. In general, the lower and middle secondary degrees yield the highest returns because they are compared to having no or only primary education. The vocational degrees and university level education yield also reasonably high returns to education.

The returns to schooling are clearly higher in Western than in Eastern states. In Western states, the returns to schooling are around 8 to 9 percentage points, while in the Eastern states, it is around 6 to 7 percentage points. Brandenburg is an exception with returns to schooling of 8 percentage points, which is above the German average. Also Lower Saxony is an exception. For this state, the lowest returns to schooling in West Germany are estimated (7.12 percentage points).

The estimated returns to schooling for each state are now related to the educational attainment in each state (see Figure 5). If wages paid for different educational attainments depend on the relative scarcity of adequately educated employees, we might expect a negative relationship between average educational attainment and individual returns to education. This correlation assumes that the demand for education is roughly equal between regions and that labour mobility is low, however.

Fig. 5 Mincerian returns to schooling vs. relative attainment in 2000

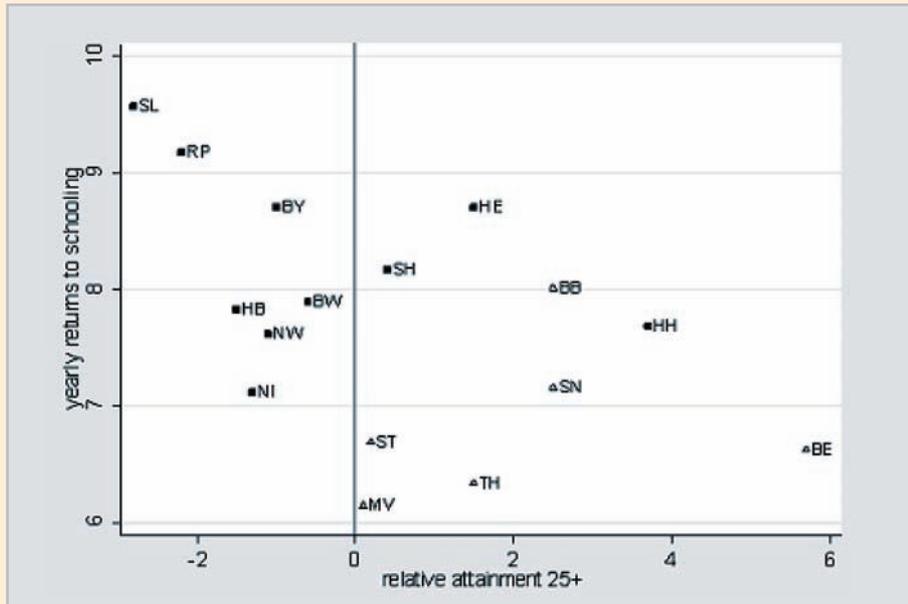
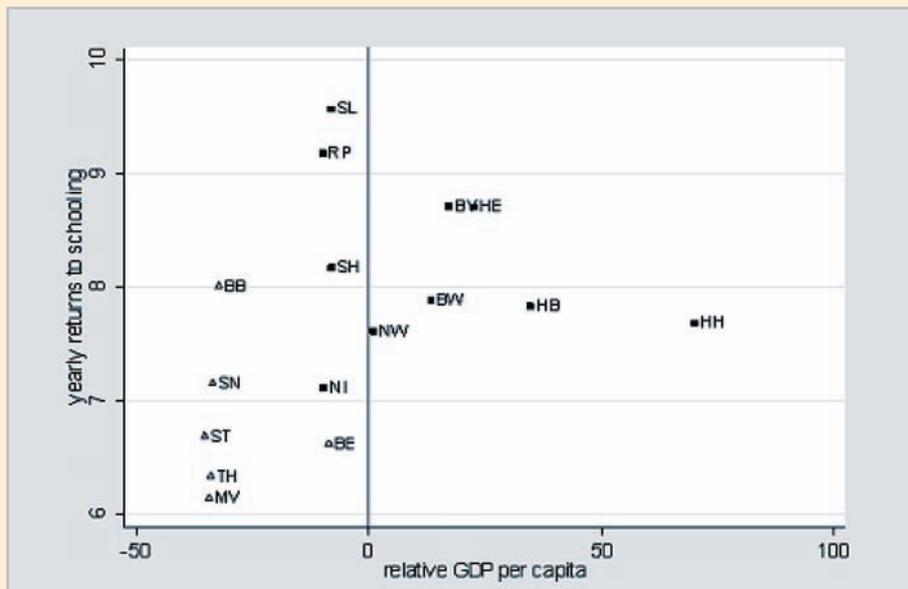


Fig. 6 Mincerian returns to schooling vs. relative GDP per capita in 2000



The relationship is not very strong but the figure suggests that there is a negative relationship between the returns to schooling and the relative educational attainment of the population. The states with the highest returns to schooling, Saarland and Rhineland-Palatinate, are also those with the lowest relative educational attainment. Hence, the evidence suggests a weak negative relation between the returns to schooling and relative educational attainment.

In a second step, we relate the estimated Mincerian returns to schooling per German state to relative GDP per capita, i.e. the deviation from the German average GDP per capita in a state (see Figure 6). The relationship between Mincerian returns to schooling and relative GDP per capita is not clear cut. If anything, the relationship is positive and either linear or concave. There is again a difference between East and West Germany. The East German states are on the lower left-hand side of the graph, hence they are the countries with relatively low GDP per capita and also low returns to schooling. Only Lower Saxony, a West German state is close to the East German states in both relative GDP per capita and estimated returns to schooling. Both city states, Hamburg and Bremen, are outliers with very high relative GDP per capita and only average returns to schooling. Saarland and Mecklenburg-Western Pomerania are outliers in the sense that returns to schooling are extremely high while their relative GDP per capita is below the German average.

5. Non-schooling variables used in the participation and employment equations		
	Participation	Employment
Sex (male)	x	x
Potential experience	x	x
Potential exp. squared	x	x
University student	x	x
Non-university student	x	x
Married	x	
Married * male	x	
Children below six	x	
Children below six * male	x	

Note: Participation includes employed and jobless persons as asked in a specific question in the micro-census.

6. Estimated contribution of a year of schooling to the probability of participation in %										
	θ	low. sec.	mid. sec.	l. sec. + app.	m. sec. + app.	sec. + voc.	h. sec.	h. sec. + app.	tech. college	univ.
Bremen	2.27	8.61	3.34	1.29	2.21	1.24	0.00	1.69	0.63	0.84
North Rh.-Westp.	2.03	8.54	1.82	1.44	2.51	1.81	0.00	1.91	0.94	1.00
Hamburg	1.74	7.19	1.17	1.54	3.15	2.11	0.00	2.75	0.93	1.08
Saarland	1.73	7.08	3.95	1.43	2.14	1.39	0.00	0.00	0.00	0.72
Lower Saxony	1.52	10.99	0.00	0.98	2.54	1.48	0.00	1.44	0.53	0.63
Berlin	1.27	7.35	1.90	0.75	2.36	2.09	0.00	0.69	0.61	0.64
Meckl.-W. Pom.	1.25	7.61	0.78	0.60	6.66	1.23	0.00	0.09	0.06	0.15
Hesse	1.18	7.23	0.66	0.99	2.54	2.32	0.00	1.61	0.83	0.73
Rhineland-Palatinate	1.17	9.34	0.32	1.21	2.21	1.75	0.00	0.87	0.36	0.60
Saxony-Anhalt	1.17	7.02	0.25	0.64	6.82	0.95	0.00	0.08	0.16	0.16
Schleswig-Holst.	0.98	6.40	3.28	1.66	1.37	1.53	0.00	0.63	0.26	0.13
Baden-Wuerttem.	0.93	8.67	0.00	0.71	1.34	1.31	0.00	0.19	0.46	0.39
Saxony	0.75	6.65	0.10	0.46	7.66	1.39	0.08	0.03	0.12	0.27
Bavaria	0.71	6.59	0.54	0.77	1.05	1.42	0.00	0.47	0.27	0.39
Brandenburg	0.70	5.90	0.85	0.62	4.70	1.05	0.00	0.13	0.15	0.16
Thuringia	0.46	6.14	0.81	0.56	5.14	1.00	0.00	0.00	0.01	0.06
Germany	1.27	8.08	0.64	0.93	2.43	1.72	0.00	1.04	0.58	0.61

Note: Estimates from participation equation using the control variables from Table 5 and data from micro-census 2000. The zero coefficients have been imposed.

Schooling, labour force participation and employment

Following de la Fuente et al. (2003), we estimate the contribution of an additional year of schooling to the probability of being employed. In the first stage we estimate a probit of labour market participation on years of schooling and the explanatory variables shown in Table 5. From this regression we construct the inverse Mills ratio, i. e. the individual probability to participate in the labour market and include it in the second stage, where we regress a dummy of employment on years of schooling and the reduced set of explanatory variables in Table 5. This approach is comparable to the two-step procedure proposed by Heckman (1979).

7. Estimated contribution of a year of schooling to the probability of being employed, conditional on labour force participation in %										
	θ	low. sec.	mid. sec.	l. sec. + app.	m. sec. + app.	sec. + voc.	h. sec.	h. sec. + app.	tech. college	univ.
Brandenburg	2.78	0.00	0.00	0.00	2.36	7.00	3.74	1.76	1.02	0.97
Saxony-Anhalt	2.59	0.00	0.00	0.00	0.00	2.74	1.90	0.66	0.00	1.12
Meckl.-W. Pom.	2.48	2.58	1.55	0.17	3.78	6.81	4.30	0.54	0.28	0.77
Bremen	2.45	0.00	7.20	3.01	1.78	4.07	0.00	2.13	1.08	2.42
Saxony	2.45	0.00	0.00	0.00	0.00	0.56	0.00	0.00	0.26	1.43
Saarland	2.39	8.57	3.55	2.94	3.57	3.96	1.14	1.35	0.35	1.15
Berlin	2.29	0.00	7.12	2.73	4.44	5.79	3.08	0.39	0.48	0.95
Thuringia	2.02	0.00	0.00	0.00	0.00	4.66	3.36	0.00	0.00	0.33
North Rh.-Westp.	1.82	6.05	2.97	1.54	2.57	3.33	0.87	1.56	1.08	0.95
Lower Saxony	1.55	6.87	1.27	1.51	2.90	4.20	1.11	1.33	1.08	0.61
Rhinel.-Palatinate	1.32	1.14	3.95	1.80	1.87	3.25	0.67	1.09	0.75	0.62
Hesse	1.30	5.50	1.26	1.33	2.68	4.08	0.86	1.29	0.99	0.77
Schleswig-Holst.	1.25	2.42	5.62	1.87	1.22	2.47	0.40	1.04	0.74	0.28
Hamburg	1.24	0.00	0.59	1.03	3.03	3.72	1.77	2.40	0.10	0.67
Baden-Wuerttem.	0.96	4.36	2.17	1.04	1.39	2.42	0.26	0.50	0.82	0.44
Bavaria	0.75	2.44	1.96	0.75	0.99	2.22	0.20	0.60	0.49	0.54
Germany	1.55	3.57	2.13	1.32	2.38	4.40	1.41	0.92	0.82	0.77

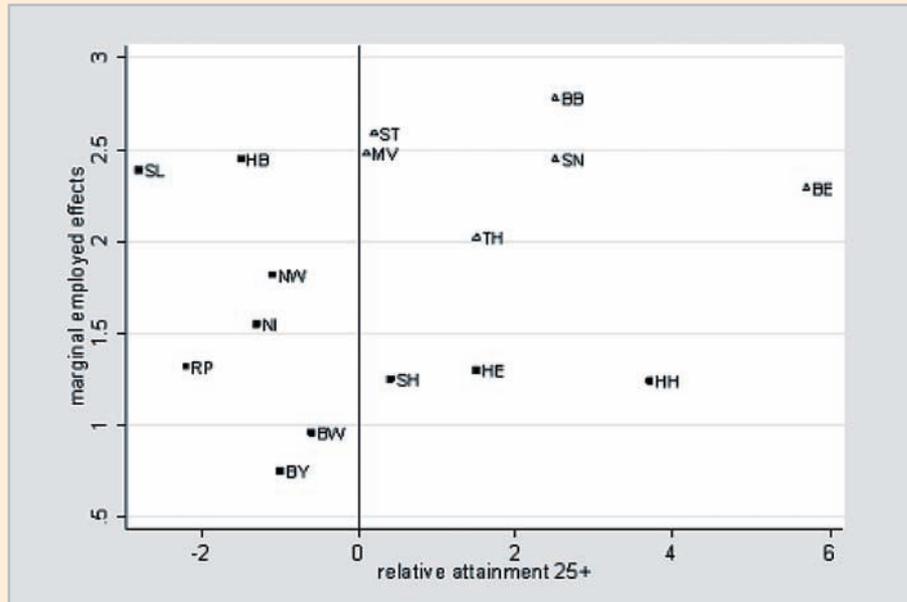
Note: Estimates from employment equation using the control variables from Table 5 and data from micro-census 2000. The zero coefficients have been imposed.

Using the micro-census 2000 and the population aged 20 – 60, the contribution of an additional year of schooling measured in percentage points for both the probability of participation and employment is shown in Table 6 and 7. Note that the estimate for employment is conditional on the probability of labour force participation, as these are the results from the second stage regression. The normal hazard function is highly significant in all regressions, i. e. the probability of labour market participation has an impact on actual employment. The estimates from a probit regression without the normal hazard function differ significantly from the conditional estimates for most states and are always higher. This could be expected because people already anticipate their chances of employment before they choose to participate in the labour market or not.

Figure 7 relates the marginal employment effect of a year of schooling to the relative educational attainment. The marginal employment effects differ widely between 0.75 and 2.78 percent. They do not seem to be directly related to educational attainment of the population in a state. Eastern states, where educational attainment is relatively high, tend to have higher marginal employment effects than Western states.

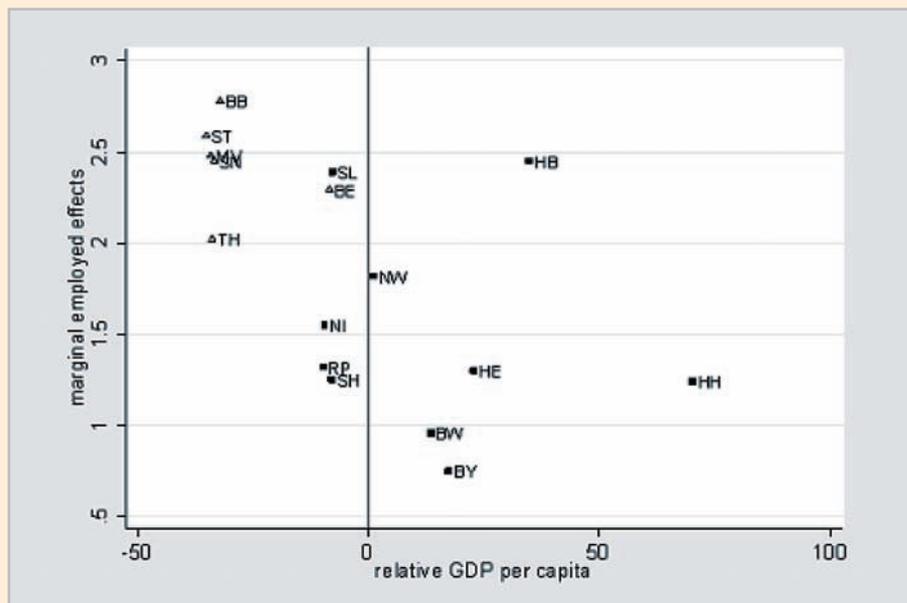
The relationship between marginal employment effects and relative GDP per capita is depicted in Figure 8. States with higher employment effects of an additional year of schooling tend to have a below-average GDP per capita and vice versa. The relationship is negative and significant. When we take out the city states of Hamburg and Bremen, which have a relatively high

Fig. 7 Marginal employment effects vs. relative attainment in 2000



GDP per capita ratio, the negative relationship is even stronger. Therefore, the significance of education for the probability of being employed rises in states with a relatively low GDP, which also tend to have higher unemployment. In states with a good economic situation like Bavaria or Baden-Wuerttemberg, education seems to be less important for finding employment.

Fig. 8 Marginal employment effects vs. GDP per capita in 2000



4 Conclusion

Our empirical analyses show the effects of public schooling expenditures on individual labour market outcomes in the German states and highlight differences between East and West German states.

Average public expenditures per student vary widely for university education. Variation in expenditures on secondary schooling is rather low. At the university level, expenditures are three times higher than expenditures on secondary schooling. Eastern states spend on average more on higher education per student than Western states, which is mainly a consequence of good (personnel) equipment and a relatively low popularity of universities and universities of applied sciences in East Germany. We therefore do not find a relationship between regional income per capita and the amount of resources available for education.

When we consider East and West Germany separately, we find a positive and significant correlation between GDP per capita and schooling attainment, however. There is a slow increase in educational attainment in Germany induced by younger cohorts being higher educated than older cohorts. Educational attainment of younger cohorts in East Germany has fallen below the national average in recent years while the regions with relatively high GDP per capita improved their position.

Schooling has an important influence on individual wages. An additional year of schooling results on average in an increase of wages by 7.8 percent. The returns to schooling are lower in East than in West Germany. We also find weak evidence of a negative correlation between yearly returns to schooling and relative schooling attainment. Therefore the tightness of the local qualification labour market seems to have an influence on the qualification wage mark-up.

The marginal employment effects of schooling are higher in East than in West German states where GDP per capita is higher and the unemployment rate is lower (mainly Baden-Württemberg, Bavaria, Hesse, and Hamburg). In these states, the qualification background seems to be less important for finding a job. Regarding wages however, the effect of education is highest in these relatively rich states. Hence, we conclude that it is essential to consider not only the impact on wages but also on employment when analysing the effect of schooling on individual labour market outcomes in different economic circumstances.

Data Appendix

1) Micro-census: Yearly representative sample of one percent of all German households. 70 percent anonymized sample for the years 1995 – 2002 provided by Statistisches Bundesamt is used here.

2) Absolute public university expenditures:
 Statistisches Bundesamt: Bildung im Zahlenspiegel 2003, page 168
 Table 8.4: Ausgaben der Hochschulen (in Mill. Euro)
 Column: Laufende Ausgaben

3) Total number of students in winter semester 2000/2001:

Statistisches Bundesamt: Bildung im Zahlenspiegel 2001, page 125

Table 7.4: Studierende im Wintersemester 2000/2001 nach Hochschularten und Ländern

Column: insgesamt

4) University graduations:

Statistisches Bundesamt: Bildung im Zahlenspiegel 2002, page 148

Table 7.14: Bestandene Prüfungen 2000 nach Prüfungsgruppe und Ländern

5) Population:

Internetdatenbank der Statistischen Ämter des Bundes und der Länder

<https://www.regionalstatistik.de/genesis/online/logon>

Sachgebiet	1	Gebiet, Bevölkerung, Erwerbstätigkeit, Wahlen
	12	Bevölkerung
	124	Bevölkerungsstand Vorausberechnung
	12411	Fortschreibung des Bevölkerungsstandes

Table 173-32-2:

Bevölkerungsstand: Durchschnittliche Jahresbevölkerung nach

Geschlecht – Jahresdurchschnitt

Column: insgesamt

6) Relative income:

Internetdatenbank der Statistischen Ämter des Bundes und der Länder

<https://www.regionalstatistik.de/genesis/online/logon>

Sachgebiet	8	Volkswirtschaftliche Gesamtrechnungen
	82	VGR der Länder
	824	Umverteilungsrechnung
	82411	Verfügbares Einkommen der privaten Haushalte

Table 666-41-2:

Verfügbares Einkommen der privaten Haushalte einschließlich

privater Organisationen ohne Erwerbszweck

Column: verfüg. Einkommen der priv. Haushalte je Einwohner

7) Graduation ratios by school type:

Internetdatenbank der Statistischen Ämter des Bundes und der Länder

<https://www.regionalstatistik.de/genesis/online/logon>

Sachgebiet	2	Bildung, Sozialleistungen; Gesundheit
	21	Bildung
	211	Allgemein bildende und berufliche Schulen
	21111	Statistik der allgemein bildenden Schulen

Table 192-61-2: Allgemein bildende Schulen: Absolventen/Abgänger

nach dem Schulabschluss – Schuljahr –

8) Employees:

Internetdatenbank der Statistischen Ämter des Bundes und der Länder

<https://www.regionalstatistik.de/genesis/online/logon>

Sachgebiet	1	Gebiet, Bevölkerung, Erwerbstätigkeit, Wahlen
	13	Erwerbstätigkeit
	133	Erwerbstätige
	13311	Erwerbstätigenrechnung des Bundes und der Länder

Table 638-32-2: Arbeitnehmer nach Wirtschaftszweigen – Jahresdurchschnitt –

Column: Arbeitnehmer im Jahresdurchschnitt – insgesamt

9) Gross domestic product & gross value added:

Internetdatenbank der Statistischen Ämter des Bundes und der Länder

<https://www.regionalstatistik.de/genesis/online/logon>

Sachgebiet	8	Volkswirtschaftliche Gesamtrechnungen
	82	VGR der Länder
	821	Entstehungsrechnung
	82111	VGR der Länder: Entstehungsrechnung

Table 426-41-2: Bruttoinlandsprodukt (BIP)/Bruttowertschöpfung (BWS)

Column: BIP zu Marktpreisen in jeweiligen Preisen

BWS zu Herstellungspreisen in jeweiligen Preisen

10) Physical Capital:

Internetquelle des Arbeitskreis VGF des Statistischen Landesamtes Baden-Württemberg:

http://www.vgrdl.de/Arbeitskreis_VGR/tab09.asp

Table: Kapitalstock in Deutschland 1991 bis 2001 nach Bundesländern

Gross stock of fixed assets (annual average) in Germany by Bundesland 1991 to 2001

11) Expenditures per pupil:

Statistisches Bundesamt

Bildung im Zahlenspiegel 2003, 2001, page 167

Table: 8.3.2. Ausgaben je Schüler sowie Personalausgaben je wöchentlicher Unterrichtsstunde nach Ländern

Column: Ausgaben je Schüller (Vollzeit und Teilzeitschüler),
allgemein bildende Schulen,

1997, 1998, 2000, 2001

For 1995 und 1996:

Statistisches Bundesamt, Mitteilungen für die Presse

Ausgaben je Schüler an öffentlichen Schulen 1995

5.März 1998

Statistisches Bundesamt, Mitteilungen für die Presse

Ausgaben je Schüler an öffentlichen Schulen 1996

23.März 1999

12) Public expenditures by state governments:

Internet source Statistisches Bundesamt

Inhalte aus der vierteljährlichen Kassenstatistik der öffentlichen Haushalte

Literature

Ammermüller, A. (2004), PISA: What makes the difference? Explaining the gap in PISA test scores between Finland and Germany, ZEW Discussion Paper No. 04 – 04, Mannheim.

Ammermüller, A. and A. Weber (2005), Educational attainment and returns to education in Germany – An analysis by subject of degree, gender and region, ZEW Discussion Paper No. 05 – 17, Mannheim.

de la Fuente, A., R. Doménech and J. Jimeno (2003), Human capital as a factor of growth and employment at the regional level. The case of Spain, Final report. http://europa.eu.int/comm/employment_social/employment_analysis/human/hum_cap_spain_regions.pdf.

Heckman, J. (1979), Sample selection bias as a specification error, *Econometrica* 47, 153 – 161.

Klös, H.-P. and R. Weiß (2003), Bildungs-Benchmarking Deutschland - Was macht ein effizientes Bildungssystem aus?, Institut der deutschen Wirtschaft, Köln.

Lauer, C. (2004), Education and labour market outcomes: A French-German comparison, ZEW Economic Studies, Physica Verlag.

Lauer, C. and V. Steiner (2000), Returns to education in West Germany – An empirical assessment, ZEW Discussion Paper No. 00 – 04, Mannheim.

Wolter, S. and B. Weber (1999), On the Measurement of Private Rates of Return to Education, *Jahrbücher für Nationalökonomie und Statistik* 218, 605 – 618.

Zwick, T. (2001), Supply of Human Capital in Times of Skill Biased Technological Change, *Jahrbücher für Nationalökonomie und Statistik* 221/3, 322 – 335.

Zwick, T. and H. Schröder (2001), Wie aktuell ist die Berufsbildung im Dienstleistungssektor?, ZEW Wirtschaftsanalysen, Bd. 55, Baden-Baden.

Autoren- und Kontaktangaben:

Andreas Ammermüller, Centre for European Economic Research (ZEW), E-Mail: ammermüller@zew.de

Anja Kuckulenz, Centre for European Economic Research (ZEW), E-Mail: kuckulenz@zew.de

PD Dr. Thomas Zwick, Centre for European Economic Research (ZEW), E-Mail: zwick@zew.de

Kein Bedarf an mehr Akademikerinnen und Akademikern

Eine Untersuchung der Einkommensentwicklung während der Ausweitung der Akademikerquote in Deutschland von 1991 bis 2001

Christiane Mück, Karen Mühlenbein

Der vorliegende Beitrag untersucht, warum trotz des beklagten Fachkräftemangels ein Hochschulstudium am deutschen Arbeitsmarkt immer weniger honoriert wird. Ein Schwerpunkt wird dabei auf den Zusammenhang zwischen der Hochschulexpansion der 1990er Jahre und der Einkommensentwicklung für Akademikerinnen und Akademiker gelegt. Die verwendete Datenbasis ist der deutsche Mikrozensus für die Jahre 1991, 1995, 1998 und 2001.

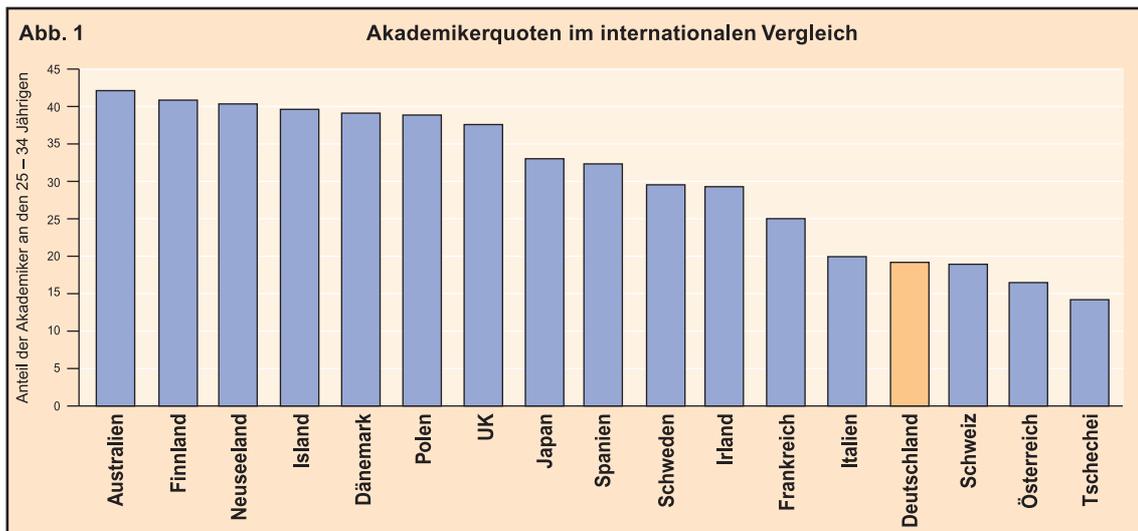
Es wird gezeigt, dass in Westdeutschland das steigende Angebot an Akademikerinnen und Akademikern auf eine konstante Nachfrage auf dem Arbeitsmarkt traf und es dadurch zu einem Preisverfall kam. Im Osten hingegen gab es zumindest zu Beginn der Untersuchungsperiode eine steigende Nachfrage und so zeigten sich trotz der Hochschulexpansion steigende Einkommensprämien für Akademikerinnen und Akademiker. Jedoch trat auch in Ostdeutschland zur Jahrtausendwende zunehmend eine Marktsättigung ein, sodass wohl zukünftig ebenfalls stagnierende oder gar abnehmende Einkommensprämien zu erwarten sind. Insgesamt lässt sich zeigen, dass Hochschulbildung zwar das Risiko von Arbeitslosigkeit verringert, ein bildungsadäquater Einkommensvorteil für die Absolventinnen und Absolventen jedoch zunehmend nicht mehr gegeben ist.

1 Ausgangssituation

Greencard-Regelung, Brain Drain, Elite-Universitäten – das alles sind zurzeit viel gebrauchte Schlagworte, die allesamt darauf hinweisen, dass es Deutschland an akademisch qualifiziertem Personal mangelt. Besonders die Unternehmen fordern eine größere Anzahl qualifizierter Akademikerinnen und Akademiker als sie der Arbeitsmarkt aktuell bereitstellen kann. Auch volkswirtschaftlich ist für Deutschland als rohstoffarmes Land das Angebot qualifizierter Arbeitskräfte eine wesentliche Voraussetzung für Wirtschaftswachstum und internationale Wettbewerbsfähigkeit.¹⁾ Verglichen mit anderen Ländern ist jedoch der Anteil an Akademikerinnen und Akademikern an der deutschen Erwerbsbevölkerung eher gering, wie Abbildung 1 zeigt.²⁾ Deutschland liegt mit 19,2 % deutlich unter dem OECD-Mittel von 31,8 %.

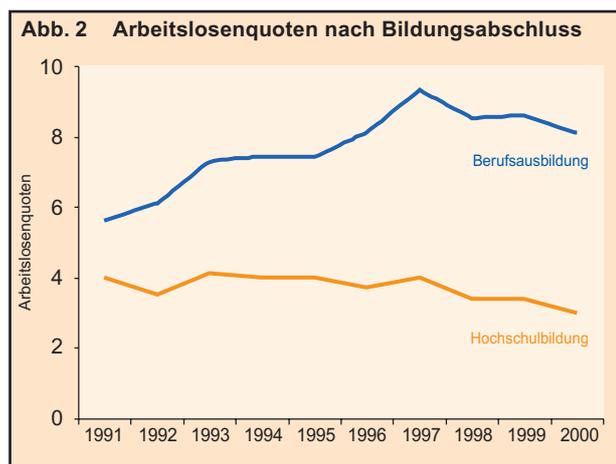
Aktuell gibt es daher in Deutschland viele Bestrebungen, den Anteil der Akademikerinnen und Akademiker unter den Erwerbstätigen zu erhöhen. Vom Hochschulsystem wird gefordert, dies durch eine Erhöhung der Absolventenzahlen zu ermöglichen. Nach Plänen der Bundesregierung sollen langfristig 40 % eines Altersjahrgangs ein Studium aufnehmen. Die ersten Maßnahmen zeigen bereits Wirkung. Im Wintersemester 2003/2004 überstieg die Studierendenzahl in Deutschland erstmals die Marke von zwei Millionen.³⁾ Es entschließen sich demnach so viele

1) Vgl. Barro/Sala-i-Martin (1995); OECD (Hsrg., 2004), S. 195ff. – 2) Vgl. OECD (Hsrg., 2004), S. 77. – 3) Vgl. http://www.destatis.de/presse/deutsch/pk/2003/ergeb_ws03_04_i.pdf



Quelle: OECD

junge Deutsche wie noch nie zu einem Studium. Von der Hochschulbildung erhoffen sich die Studienanfänger/-innen im späteren Berufsleben geringere Arbeitslosigkeit und ein hohes Einkommen.⁴⁾



Quelle: Institut für Arbeitsmarkt und Berufsforschung

Der vorliegende Beitrag untersucht, in wie weit diese Erwartungen auch in Zeiten der Hochschulexpansion erfüllt werden. Erste Indikatoren zeigen ein differenziertes Bild. Wie Abbildung 2 darstellt, ist Hochschulbildung trotz steigender Akademikerzahlen in Deutschland immer noch ein guter Schutz vor Arbeitslosigkeit. Die monetäre Wertschätzung akademischer Qualifikation im deutschen Arbeitsmarkt, gemessen am Einkommensdifferenzial von Akademikerinnen und Akademikern im Vergleich zu anderen Bildungsgruppen, geht jedoch zurück.

Das Einkommensdifferenzial ist die absolute Differenz zwischen dem durchschnittlichen Akademikereinkommen und dem Durchschnittseinkommen eines Nicht-Akademikers, die Einkommensprämie der entsprechende prozentuale Anteil. Während eine Akademikerin bzw. ein Akademiker im Jahre 1992 noch 70 % bzw. 75% mehr verdiente als eine Absolventin bzw. ein Absolvent der Sekundarstufe II, betrug die Einkommensprämie 2002 nur noch 56 % bzw. 57 %.⁵⁾

Trotz aller Beteuerungen eines steigenden Bedarfs an Akademikerinnen und Akademikern ist also ein Hochschulstudium auf dem Arbeitsmarkt immer weniger wert. Wenn die Absolventinnen und Absolventen auf dem Arbeitsmarkt jedoch keine adäquate Rendite für ihr langes Studium erwarten können, weshalb sollten sie überhaupt noch studieren? Lohnt sich ein Studium im Vergleich zu anderen Ausbildungsformen noch?

4) Vgl. OECD (Hrsg., 2004), S. 165, 176. – 5) Vgl. OECD (Hrsg., 1995), S. 232f.; OECD (Hrsg., 2004), S. 187.

In dieser Untersuchung soll vor dem Hintergrund einer fallenden Einkommensprämie für Akademikerinnen und Akademiker detailliert analysiert werden, weshalb trotz des beklagten Fachkräftemangels ein Hochschulstudium am deutschen Arbeitsmarkt immer weniger honoriert wird. Zunächst werden hierfür kurz Erklärungsansätze aus der bisherigen Forschung vorgestellt und ihre Ergebnisse diskutiert. Ein Schwerpunkt wird dabei auf den Zusammenhang zwischen Einkommensentwicklung und akademischer Bildungsbeteiligung gelegt. Anschließend sollen mögliche Ursachen für den beobachteten Rückgang der Einkommensprämie zwischen 1991 und 2001 empirisch analysiert werden. Zuerst soll geprüft werden, ob die Nachfrage nach höher qualifizierten Arbeitnehmerinnen und Arbeitnehmern in gleichem Maße angestiegen ist wie das Angebot. Als zweites wird aufgezeigt, in wie weit die Bildungsexpansion zu Veränderungen in der Verteilungsstruktur der Arbeitseinkommen geführt hat. Hierfür werden sowohl die Einkommensprämien unterschiedlicher Alters- bzw. Erfahrungskohorten als auch die aktuelle Streuung der Einkommensprämien über alle Erfahrungskohorten betrachtet.

2 Stand der Literatur

Renditen in Abhängigkeit vom Bildungsabschluss

Nach der in den 1960er Jahren entwickelten Humankapitaltheorie kann die Anhäufung von Humankapital durch (Hochschul-)Bildung als eine Investition betrachtet werden, die aufgrund des Verzichts auf Einkommen während der Zeit der Ausbildung zunächst Kosten verursacht, später aber mit einem höheren Einkommen verbunden ist.⁶⁾ Ökonomisch rational handelnde Individuen werden in Humankapital investieren, wenn die erwarteten Erträge die Kosten übersteigen. Da die Gesamtkosten eines Studiums jedoch nur schwer abschätzbar sind, werden sich rationale Individuen bei zu geringen erwarteten Erträgen des Studiums nach dem Ende ihrer Schullaufbahn eventuell gegen ein Studium entscheiden, obwohl sie dafür ausreichend begabt wären.

In Folge der Bildungsexpansion der 1970er Jahre wurden in Deutschland eine Reihe von Studien zu den Wechselwirkungen zwischen Bildungs- und Beschäftigungssystem erstellt. Hierbei wurden sowohl theoretisch als auch empirisch die veränderten Berufs- und Arbeitsmarktchancen von Akademikerinnen und Akademikern untersucht.⁷⁾ Beispielsweise war Gegenstand der Forschung, ob die Erhöhung der Bildungsbeteiligung zu erhöhter Erwerbstätigkeit führte.⁸⁾

Aktuelle deutsche Studien befassen sich primär mit Bildungsrenditen oder Aspekten der Verteilungsstruktur bei Bildung und Einkommen. Hierbei liegt der Schwerpunkt auf der Betrachtung der Einflüsse auf die Höhe von Bildungsrenditen⁹⁾ und auf dem Vergleich mit anderen Ländern, besonders in Europa und Nordamerika.¹⁰⁾ Auch werden die Einkommensverläufe unterschiedlicher Bildungsgruppen untersucht.¹¹⁾

6) Vgl. Mincer (1997) für neuere Zusammenfassungen der damaligen Diskussion. Die Humankapitaltheorie in klassischer Form ignoriert alle weiteren positiven, auch nicht-monetären Effekte, die mit einem höheren Bildungsstand einhergehen. Dieses Vorgehen wird für die vorliegende Arbeit übernommen. – 7) Vgl. Blossfeld (1983). – 8) Vgl. Blien/Reinberg/Tessaring (1990); Reinberg/Fischer/Tessaring (1995). – 9) Vgl. Steiner/Lauer (2000); Ammermüller/Dohmen (2004). – 10) Vgl. Brauns/Müller/Steinmann (1997). – 11) Vgl. Fitzenberger/Garloff/Kohn (2003).

Renditen bei einer Ausweitung der Bildungsbeteiligung

Während der Humankapital-Ansatz von einer „ceteris paribus“-Betrachtung beim positiven Zusammenhang zwischen Bildungsgrad und Bildungsrendite ausgeht, könnte sich bei einer gleichzeitigen Ausweitung des Bildungsniveaus der Bevölkerung (z. B. durch eine Öffnung des Bildungssystems oder durch geburtenstarke Jahrgänge) eine Situation ergeben, bei der trotz eines höheren Bildungsstandes aufgrund eines Überangebots Höhergebildeter keine hohen Bildungsrenditen erzielt werden können. Ein solcher Zusammenhang galt offenbar in den 1980er und 1990er Jahren in Österreich und Deutschland.

Fersterer/Winter-Ebmer (1999) weisen in Österreich im Zeitraum von 1981 bis 1997 sinkende Bildungsrenditen für Hochschulbildung bei einer gleichzeitigen Ausweitung der Bildungsbeteiligung nach. Während der Anteil der männlichen Bevölkerung mit Hochschulabschluss um 30 % und der an der weiblichen Bevölkerung um über 85 % anstieg, sanken die jeweiligen Bildungsrenditen von 10 % auf 7,4 % bzw. von 11,4 % auf 8 %. Gleichzeitig hat die Streuung der Bildungsrenditen zugenommen.

Steiner/Lauer (2000) untersuchen die Veränderung der Bildungsrenditen in Deutschland zwischen 1984 und 1997 auf Basis eines Anstiegs an Hochschulabsolvent(inn)en. Die Rendite eines zusätzlichen Jahres Hochschulbildung betrug durchschnittlich 10,5 % für Frauen und 8,3 % für Männer. Allerdings konnte, besonders bei jüngeren Hochschulabsolvent(inn)en, ein Absinken der Bildungsrenditen nachgewiesen werden. Zudem sind die Renditen zusätzlicher Bildungsjahre für Fachhochschulabsolvent(inn)en durchwegs höher als für Absolvent(inn)en von Universitäten.

Bei ausländischen Studien hingegen, vor allem in den USA, steht der Zusammenhang zwischen Bildungsexpansion, technologischem Fortschritt und Einkommenswachstum im Vordergrund. Viele dieser empirischen Untersuchungen zeigen einen positiven Zusammenhang zwischen der Erhöhung des Bildungsstands und den Einkommensprämien.¹²⁾ Qualifikations-komplementäre technische Veränderungen (skill-biased technology change) steigern die Nachfrage nach Qualifikationen, so dass das höhere Angebot an qualifizierten Arbeitskräften mehr als kompensiert wird. Dies führt selbst bei einer Erhöhung der Bildungsbeteiligung zu steigenden Renditen eines Studiums.¹³⁾

3 Forschungslücke

Für Deutschland gibt es – mit Ausnahme von Steiner/Lauer (2000) – aktuell keine Studie, welche den Einfluss der Ausweitung universitärer Bildungsbeteiligung während der 1990er Jahre auf den Arbeitsmarkt untersucht. Besonders wichtig ist hierbei, ob sich die Wertschätzung des Arbeitsmarkts für Hochschulbildung – ausgedrückt im Einkommensdifferenzial zu anderen Ausbildungsformen – verändert hat. Zudem werden vielfach die Ursachen für das Einkommensdifferenzial und für dessen Veränderung über die Zeit nur am Rande untersucht. Besonders eine Untersuchung zu einer Erklärung der Veränderung des Einkommensdifferenzials anhand des Zusammenspiels von Angebot und Nachfrage fehlt bislang.

¹²⁾ Zum Beispiel Katz/Murphy (1992), die eine Erhöhung der Bildungsrendite von Akademikern im Vergleich zu High-School-Absolventen von 61 % auf 82 % zwischen 1961 und 1979 nachweisen. – ¹³⁾ Vgl. Katz/Murphy (1992); Card (2001); Card/Lemieux (2001); Acemoglu (2003a); Acemoglu (2003b).

4 Forschungsfragen

Vor dem Hintergrund des oben beschriebenen Forschungsstands und unter der Annahme eines funktionierenden Arbeitsmarkts, der für Hochschulbildung zu positiven Einkommensprämien führt, untersuchen wir drei Fragestellungen, die miteinander in Zusammenhang stehen.

Hat die steigende Akademikerquote dazu geführt, dass die Einkommensprämie für ein Hochschulstudium gesunken ist? Dies würde dafür sprechen, dass bei einem funktionierenden Marktmechanismus ein Überangebot von Akademikerinnen und Akademikern auf dem Arbeitsmarkt besteht und dadurch die „Preise“ für Akademikerinnen und Akademikern sinken.

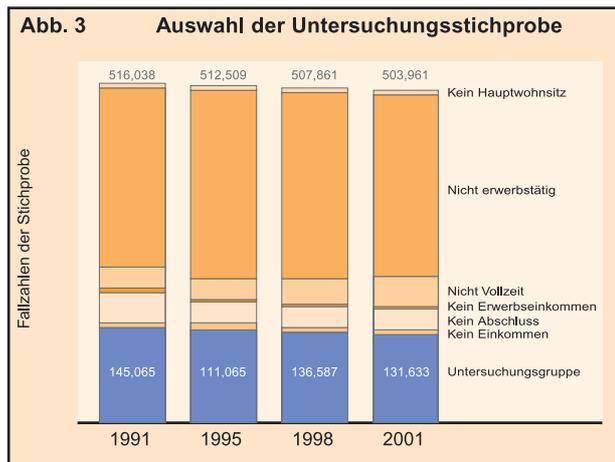
Sind die Gehälter aller Altersgruppen von Hochschulabsolvent(inn)en gleich stark gesunken? Wenn es einen stärkeren Gehaltsverlust bei den Absolventinnen und Absolventen gab, die nach der Ausweitung der Akademikerquote ihr Studium beendet haben, ist davon auszugehen, dass die Unternehmen die höhere Zahl von Absolvent(inn)en als Qualitätsverlust der Hochschulausbildung interpretiert haben.

Hat die Streuung der Gehälter von Akademikerinnen und Akademikern nach der Ausweitung der Akademikerquote zugenommen? Dies könnte dadurch begründet sein, dass Arbeitgeber zwar für eine bestimmte Anzahl von Absolventinnen und Absolventen nach wie vor hoch bezahlte Posten anbieten, für andere jedoch nur schlechter bezahlte Arbeitsstellen übrig bleiben.

5 Datengrundlage und Klassifikation

Grundlage der Untersuchung sind die Daten aus dem Mikrozensus der Bundesrepublik Deutschland. Der Mikrozensus eignet sich vor allem durch die Größe der Stichprobe sehr gut für eine derartige Untersuchung. Selbst wenn nur eine bestimmte Bevölkerungsgruppe (in unserem Fall Vollzeitberufstätige) analysiert wird und vielfältige Subgruppen gebildet werden, stehen noch genügend Datenpunkte zur Verfügung, um statistisch signifikante und repräsentative Aussagen zu ermöglichen. Auch die Erhebungsmethode spricht für die Verwendung der Daten des Mikrozensus. Die Befragten werden mittels einer einstufig geschichteten Flächenstichprobe zufällig ausgewählt und die erhobenen Tatbestände sind überwiegend mit einer Auskunftspflicht belegt. Daher werden durch die Natur der Datenerhebung Verzerrungsproblematiken weitgehend ausgeschlossen. Besonders im Hinblick auf die von uns verwendeten Einkommensdaten ist dies von großer Bedeutung.

Dieser Beitrag basiert auf den Mikrozensus Daten aus den Jahren 1991, 1995, 1998 und 2001. Im Jahr 1991 wurden erstmals detaillierte Bildungsmerkmale (letzter Schulabschluss, letzter beruflicher Ausbildungsabschluss) erhoben. Allerdings wurden diese Angaben zuerst nur alle zwei Jahre erfragt. Seit 1998 gehören diese Angaben jedoch zum jährlichen Grundprogramm des Mikrozensus. Die Auswahl der untersuchten Jahre ergibt sich daher vor allem aus der Verfügbarkeit der relevanten Daten.



Quelle: Eigene Analyse

Nach diesen Einschränkungen bleiben für die einzelnen Jahre noch 145 065 (1991), 141 065 (1995), 136 587 (1998) bzw. 131 633 (2001) Datensätze übrig. Diese stellen die Auswertungsbasis für unsere Arbeit dar (vgl. Abbildung 3).

Die auszuwertende Datenmenge wird zudem in verschiedene Gruppen nach Geschlecht und Ost- bzw. Westdeutschland unterteilt. Die Aufteilung anhand der formalen Qualifikationsstruktur erfolgt durch zwei unterschiedliche Klassen von beruflichen Bildungsabschlüssen:¹⁴⁾

1) Lehr-/Anlernausbildung, berufliches Praktikum oder Berufsfachschulabschluss, Meister/Techniker oder Fachschulabschluss, Fachschule der DDR und (2) Fachhochschulabschluss, Universitätsabschluss, Promotion.

Insgesamt erhalten wir damit einen Datensatz mit 32 verschiedenen Gruppen von Arbeitnehmern (zwei Bildungsabschlüsse, vier Betrachtungszeitpunkte, Ost- und Westdeutschland, Geschlecht).¹⁵⁾ Für jede dieser Gruppen wird das durchschnittliche Einkommen berechnet. Die verschiedenen mittleren Einkommen dienen als Grundlage der folgenden Untersuchungen.

Ergänzend zur Aufteilung nach Herkunft, Geschlecht und Ausbildungsabschluss bilden wir sechs Kohorten, die für die unterschiedliche Länge der Berufserfahrung Rechnung tragen. Das tatsächliche Jahr des Bildungsabschlusses und damit der Eintrittszeitpunkt in das Beschäftigungssystem wird erst seit 2001 im Mikrozensus erhoben. Daher treffen wir für die anderen Betrachtungszeitpunkte die Annahme, dass das Durchschnittsalter für die Absolvent(inn)en der Klasse (1) 23 Jahre beträgt, während die Absolvent(inn)en der Klasse (2) mit durchschnittlich 28 Jahre ihre Ausbildung beenden. Der Berufseinstieg erfolgt nach unserer Definition jeweils im Lebensjahr nach Vollendung der Berufsausbildung. Auf Basis des Lebensalters und der angenommenen Ausbildungsdauer errechnen wir die Berufserfahrung der Einkommensempfänger/-innen als Lebensalter abzüglich des geschätzten Alters bei Berufseintritt (vgl. Tabelle 1).¹⁶⁾

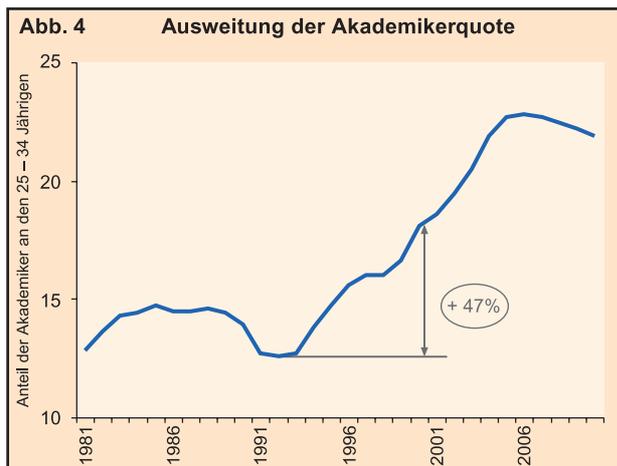
14) Im Gegensatz zu anderen Studien verwenden wir als Vergleichsgruppe für die Hochschulabsolvent(inn)en nicht die Schulabgänger/-innen ohne Berufsausbildung, sondern die Absolvent(inn)en einer Berufsausbildung. Hierdurch wird die Entscheidungssituation einer Schulabgängerin bzw. eines Schulabgängers wesentlich besser repräsentiert. – 15) Für ein ähnliches Vorgehen vgl. Katz/Murphy (1992) sowie Fersterer/Winter-Ebmer (1999). – 16) Vgl. Steiner/Lauer (2001).

1. Einteilung nach Erfahrungskohorten		
	Berufsausbildung	Hochschulabsolvent(inn)en
Kohorte nach Berufserfahrung	Durchschnittliches Alter bei Abschluss: 23 Jahre	Durchschnittliches Alter bei Abschluss: 28 Jahre
Kohorte I: < 5 Jahre	< 28 Jahre	< 32 Jahre
Kohorte II: 5 – 10 Jahre	28 – 32 Jahre	33 – 37 Jahre
Kohorte III: 10 – 15 Jahre	33 – 37 Jahre	38 – 42 Jahre
Kohorte IV: 15 – 20 Jahre	38 – 42 Jahre	43 – 47 Jahre
Kohorte V: 20 – 25 Jahre	43 – 47 Jahre	48 – 52 Jahre
Kohorte VI: > 25 Jahre	> 48 Jahre	> 53 Jahre

Durch diese Kohortenperspektive lassen sich Berufsanfänger/-innen von erfahrenen Arbeitnehmern unterscheiden. Diese Unterscheidung ist wichtig, da sich „der wesentliche Wandel der Berufsstruktur mit dem Eintritt junger Arbeitskräfte in das Beschäftigungssystem vollzieht. [...] Die Analyse der Wirkungen des Ereignisses Bildungsexpansion muss sich daher auf den Eintrittszeitpunkt der Erwerbstätigen in das Beschäftigungssystem beziehen.“¹⁷⁾

6 Ergebnisse

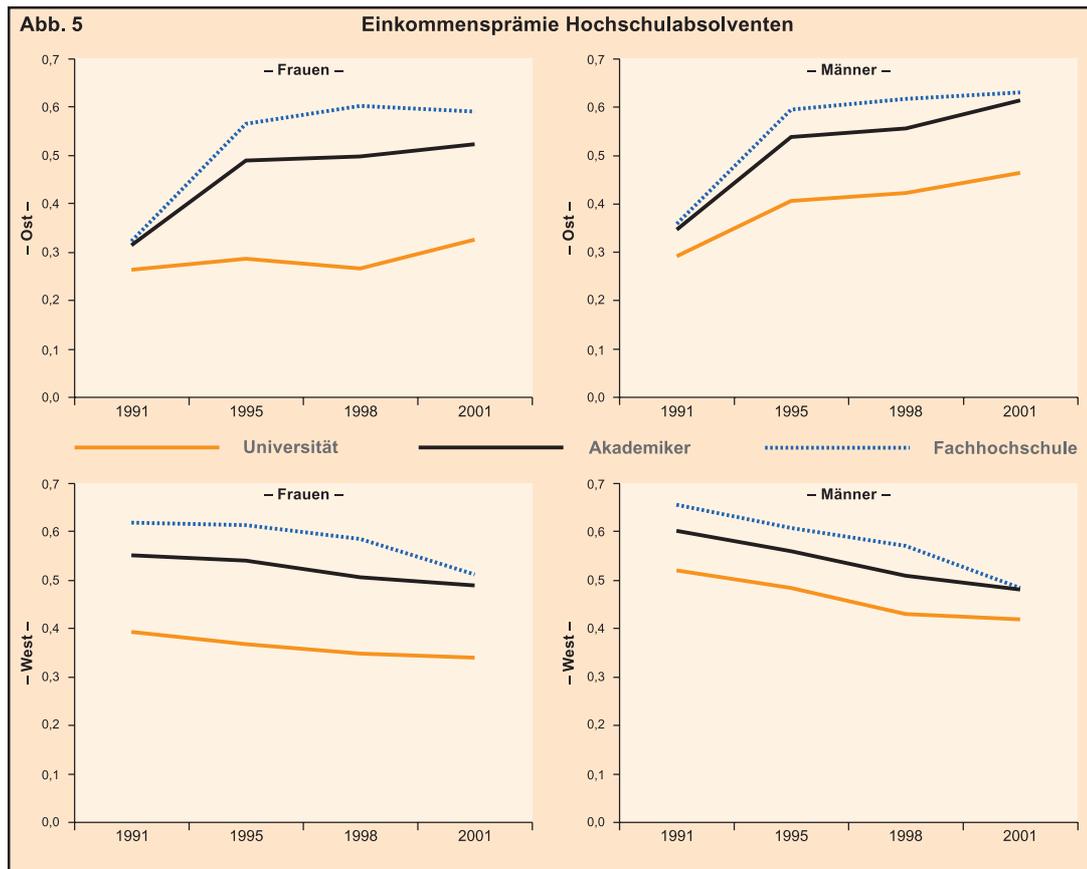
Einkommensprämien



Die einfachste Erklärung für das Absinken der Einkommensprämie für ein Hochschulstudium ist, dass die Bildungsexpansion zu einem Überangebot an Akademikerinnen und Akademikern geführt hat. Deutschland verzeichnete im untersuchten Zeitraum einen starken Anstieg der Quote der Hochschulabsolvent(inn)en (vgl. Abbildung 4). Während 1991 weniger als 13 % der 25 – 32-Jährigen einen Hochschulabschluss vorweisen konnten, waren es 2001 bereits knapp 19 %. Die Hochschulabsolventenquote ist also in 10 Jahren um fast 50 % gestiegen.

Interessant ist nun zu sehen, welche Auswirkungen diese Mengenausweitung auf dem Arbeitsmarkt hatte. Als Reaktion auf das starke Mengenwachstum stellte sich im Westen eine Senkung der Einkommensprämien in Höhe von 8 % bzw. 15 % für Akademikerinnen und Akademiker ein (vgl. Abbildung 5). Die Mengenausweitung hatte somit einen negativen Preiseffekt, die Einkommensprämien sind jedoch noch deutlich positiv. Im Osten hingegen stiegen die Einkommensprämie trotz des Anstiegs der Akademikerquote um 58 % bzw. 60 % für Akademikerinnen und Akademiker. Der Arbeitsmarkt im Osten hat demnach die höhere Zahl von Akademikerinnen und Akademikern aufgenommen und hat seine Wertschätzung durch höhere Einkommensprämien ausgedrückt.

¹⁷⁾ Blossfeld (1983), S. 207.



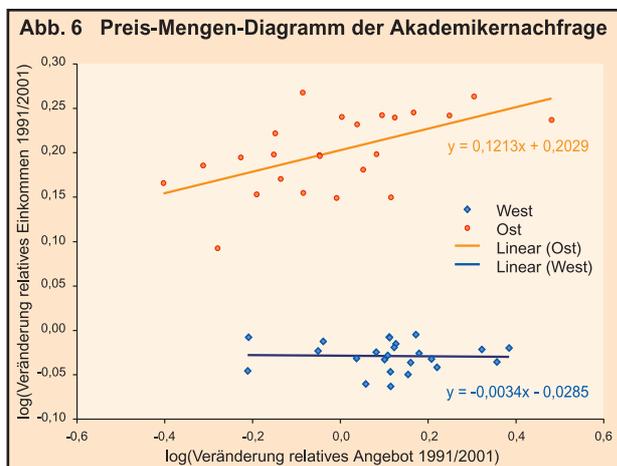
Um zu zeigen, wie Mengenzuwachs und Einkommensveränderung zusammenhängen, berechnen wir für jede unserer Datengruppen sowohl die Veränderungen der Angebotsmenge als auch der Einkommensprämien für den Zeitraum zwischen 1991 und 2001. Die logarithmierte relative Veränderung im Angebot ($\log \Delta s$) ist die Angebotssteigerung in einer bestimmten Datengruppe. Die logarithmierte relative Veränderung des Durchschnittseinkommens wird in Form der Einkommensprämie der Absolventinnen und Absolventen eines Hochschulstudiums gegenüber den Absolventinnen und Absolventen einer Berufsausbildung dargestellt ($\log \Delta w$). Die 48 Datenpunkte (Geschlecht, Ost- oder Westdeutschland, 6 Erfahrungskohorten) für die Veränderung zwischen 1991 und 2001 werden in ein Mengen-Preis-Diagramm in Abbildung 6 eingetragen. Die Trendlinien ergeben sich aus den Regressionen.

Diese einfache Methode gibt uns einen ersten Anhaltspunkt zur relativen Entwicklung von Angebot und Nachfrage nach einzelnen Qualifikationsgruppen. Das Ergebnis ist in Abbildung 6 dargestellt. Wir beobachten einen statistisch signifikanten Zusammenhang zwischen den Veränderungen im Angebot und den Veränderungen der Einkommensprämien jedoch nur, wenn wir die Betrachtung in Ost- und Westdeutschland unterteilen.

In Ostdeutschland hat die Ausweitung der Bildungsbeteiligung zu einem starken Anstieg der Einkommensprämien für Hochschulabsolvent(inn)en geführt. Hieraus ist zu erkennen, dass ein Anstieg der Akademikerquote zu einer deutlichen Einkommenssteigerung führte. Der Arbeitsmarkt hat die zusätzlichen Absolvent(inn)en demnach nicht nur aufgenommen, sondern sieht sie immer noch als so knappes Gut an, dass sich die Einkommen im Durchschnitt erhöht haben. Wie in den

USA führt die Ausweitung der Bildungsbeteiligung zu Steigerungen in der Einkommensprämie. Der Bedarf an Akademikerinnen und Akademikern ist schneller gewachsen als das Angebot, so dass deren Einkommen stark angestiegen sind.

In Westdeutschland ist die Situation dagegen anders: Der Anstieg der Akademikerquoten hatte einen negativen Effekt auf die Höhe der Einkommensprämie, was bedeutet, dass eine Ausweitung der akademischen Bildungsbeteiligung zu einer Verringerung der Einkommensprämie geführt hat. Die Ergebnisse für den Westen Deutschlands stimmen im Wesentlichen mit denen der Untersuchungen von Festerer/Winter-Ebmer (1999) für Österreich und von Steiner/Lauer (2000) für Deutschland überein. Demnach hat die Ausweitung der Akademikerquote in Westdeutschland nicht zu qualifikations-komplementären technischen Veränderungen geführt, aufgrund derer Hochschulbildung in noch stärkerem Maß nachgefragt wurde. Es zeigt sich vielmehr ein relativ starrer Bedarf an Akademikerinnen und Akademikern: Eine Markträumung (bei gleich bleibender qualifikationsspezifischer Arbeitslosenquote) tritt nur dann ein, wenn die Einkommensprämien im Vergleich zu niedriger qualifizierten Tätigkeiten sinken.



Diese Untersuchung zeigt, dass sich Veränderungen im Einkommen zu einem großen Teil aus Veränderungen im Angebot erklären lassen. Für Ostdeutschland beträgt die durch die Trendlinie erklärte Varianz der Einkommenssteigerung 23 %, im Westen sogar knapp 50 %. Ein direkter Wirkungszusammenhang ist dadurch jedoch nicht abzuleiten, denn die unterschiedlichen Auswirkungen einer erhöhten Akademikerquote in Ost und West legen nahe, dass es weitere Faktoren gibt, welche die Aufnahme der Akademikerinnen und Akademiker zu adä-

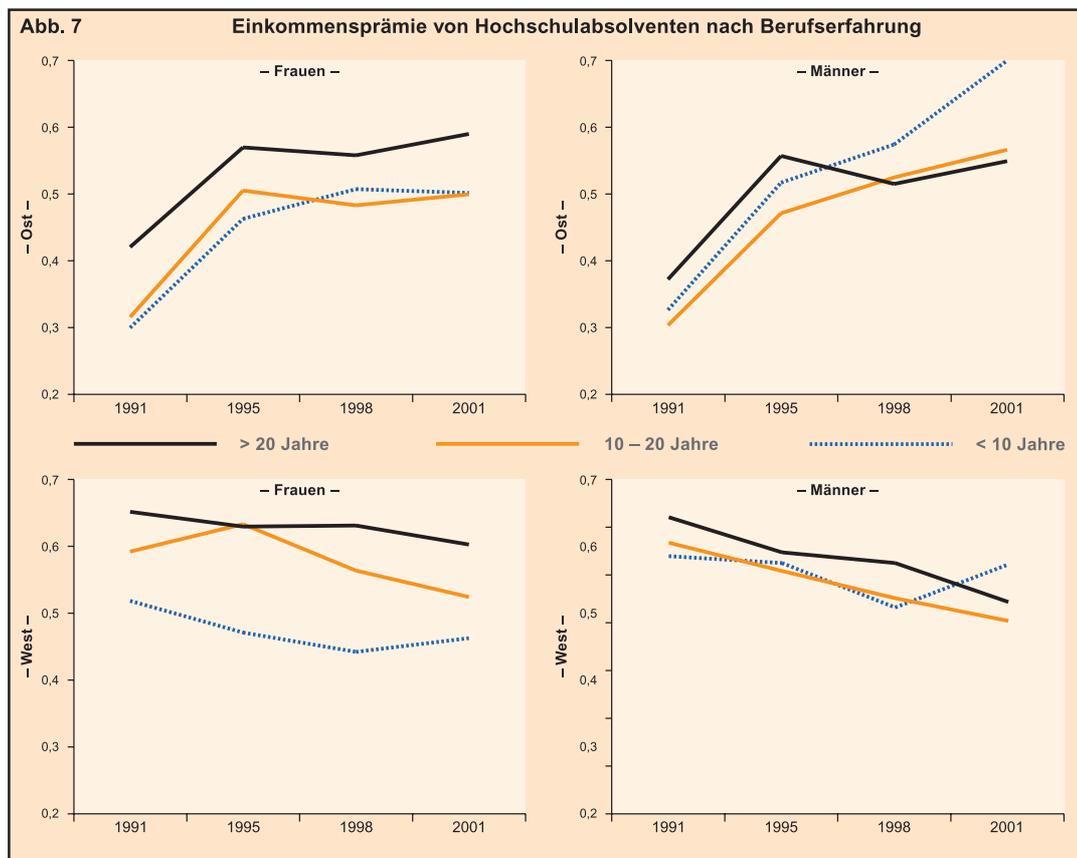
quaten Einkommen im Arbeitsmarkt beeinflussen. Dies können beispielsweise absolute Unterschiede in der Akademikerquote, Strukturunterschiede der Industrien und im Arbeitsmarkt sein.

7 Qualitätseffekte

Zusätzlich zu dem reinen Mengeneffekt kann auch ein Absinken der durchschnittlichen Absolventenqualität zu einer Verringerung des durchschnittlichen Einkommens führen. So kann es beispielsweise sein, dass durch die Ausweitung der akademischen Bildungsbeteiligung das durchschnittliche persönliche Fähigkeitsniveau der Absolvent(inn)en abgesunken ist. Gleichzeitig kann auch eine Überlastung der Hochschulen zu einem Absinken der dort erworbenen Qualifikation führen.

Diese möglichen Qualitäts- oder Mengeneffekte einer Ausweitung der Hochschulabsolvent(inn)enquote kann man empirisch differenzieren. Die zugrunde liegenden Annahmen sind sehr simpel: Ein reiner Mengeneffekt würde zu einem Absinken der Einkommensprämie von

Hochschulabsolvent(inn)en über alle Alterskohorten hinweg führen. Das heißt, dass die Einkommensprämie akademisch qualifizierter Berufsanfänger/-innen zwischen 1991 und 2001 genauso stark gefallen sein müsste wie die Einkommensprämie der Kohorten mit Berufserfahrung. Ein Qualitätseffekt, also eine Verschlechterung des Ausbildungsergebnisses aufgrund individueller oder institutioneller Ursachen, dürfte sich jedoch nur bei Akademikerinnen und Akademikern zeigen, die ihren Abschluss nach der Bildungsexpansion erworben haben; ältere Hochschulabsolvent(inn)en wären nicht davon betroffen. Mittels des „difference in difference“ – Ansatzes, der die unterschiedlichen Entwicklungen der Einkommensprämien je nach Berufserfahrung berücksichtigt, kann dargestellt werden, welcher der beiden Effekte vorliegt.¹⁸⁾ Abbildung 7 gibt einen Überblick über die Entwicklung der Einkommensprämien der Kohorten an, aufgeteilt nach Geschlecht und Herkunft.



Hinweise auf einen Qualitätseffekt durch eine Ausweitung der Bildungsbeteiligung liefern vor allem diejenigen Fälle, in denen sich die Trendlinien der Einkommensprämien für unerfahrene (weniger als 10 Jahre Berufserfahrung) und sehr erfahrene (mehr als 20 Jahre Berufserfahrung) Arbeitnehmer in unterschiedliche Richtungen bewegen.

Im Westen ist dies vor allem bei den Frauen der Fall: Während erfahrene Akademikerinnen hier eine nahezu konstant hohe Einkommensprämie erhalten, fällt die Prämie der jungen Hochschul-

18) Vgl. Angrist/Krueger (1998), S. 21ff., Lui/Suen (2003). Wir stellen hier nur eine „difference“ dar, nämlich die Veränderung der Einkommensprämien innerhalb einer Gruppe (Ost, West; Frauen, Männer). Die zweite „difference“, die Veränderung über die Jahre hinweg innerhalb einer Gruppe, ist aus Abbildung 7 zu entnehmen.

absolventinnen. Wenn die Einkommensprämie zumindest teilweise ein Indiz für die wahrgenommene Qualität der Ausbildung ist, ist folglich aus Sicht des Arbeitsmarkts die Ausbildungsqualität von Akademikerinnen gesunken. Dies kann zum Teil auf den überdurchschnittlich hohen Anteil von Studentinnen am allgemeinen Wachstum der akademischen Bildungsbeteiligung zurück zu führen sein. Bei den Männern im Westen ist sowohl bei erfahrenen als auch bei unerfahrenen Akademikern ein Absinken der Einkommensprämie zu beobachten. Hier greift demnach hauptsächlich der Mengeneffekt.

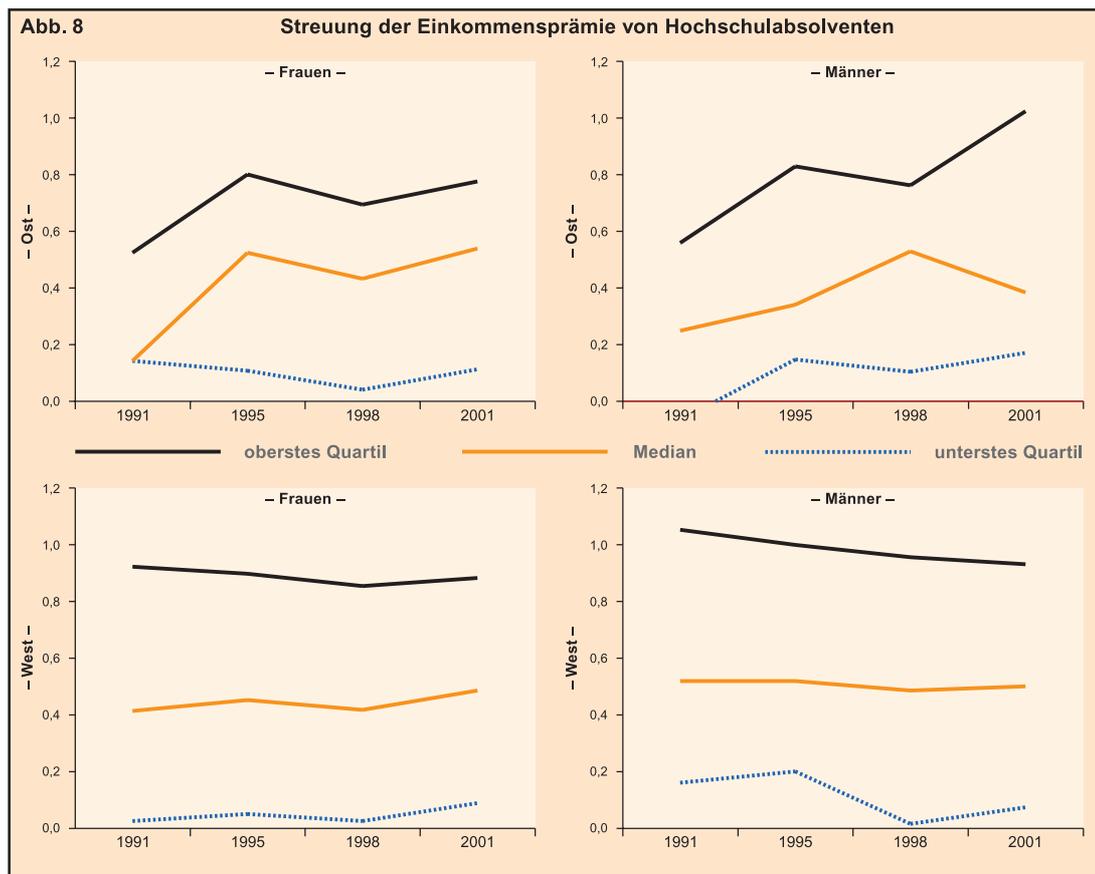
Im Osten dagegen ist zu beobachten, dass sowohl bei Männern als auch bei Frauen die unerfahrenen Hochschulabsolvent(inn)en höhere Wachstumsraten bei den Einkommensprämien erreichen als die erfahrenen Akademikerinnen und Akademiker. Bei den Männern im Osten liegt die Einkommensprämie der unerfahrenen akademischen Berufsanfänger sogar über der für Akademiker mit mehr als 20 Jahren Berufserfahrung. Demnach wird im Osten das aktuelle Fachwissen junger Hochschulabsolventen wohl sehr hoch geschätzt. Dies würde der These entsprechen, dass der Grund für den Anstieg der Einkommensprämien im Osten eine qualifikations- – komplementäre technische Entwicklung ist. Hier wären junge Akademikerinnen und Akademiker im Vorteil, die mit den aktuellen Konzepten und Methoden besser vertraut sind als ihre berufserfahreneren Kolleginnen und Kollegen.

Insgesamt kommt es jedoch über die Zeit, besonders bei den unerfahrenen Akademikerinnen und Akademikern, zu einer Angleichung der Einkommensprämien in Ost- und Westdeutschland. Da die Einkommensprämie in Relation zu den Absolvent(inn)en einer Berufsausbildung in der jeweiligen Referenzgruppe (Ost – West, Männer – Frauen) berechnet wird, kann hierdurch nicht auf eine Angleichung der Gehälter in Ost und West geschlossen werden.

8 Streuung der Einkommen

Bei dem beobachteten generellen Verfall der Bildungsrenditen wäre es möglich, dass Einkommensprämien in unterschiedlicher Höhe in unterschiedlichem Ausmaß betroffen wären. So wäre es denkbar, dass vor allem geringe Einkommensprämien gefallen sind, während hohe Einkommensprämien nahezu konstant blieben. In diesem Fall hätte sich die Streuung der Einkommensprämien erhöht. Ein solches Phänomen ist oft dann zu beobachten, wenn es einen relativ geschlossenen Arbeitsmarkt für hochbezahlte Akademikerjobs gibt, der von einem allgemeinen Absinken der Einkommensprämien nahezu unberührt bleibt. Zur Beantwortung dieser Frage haben wir die Entwicklung der Einkommensprämien an den Quartilsgrenzen zu den untersten 25 %, 50 % und 75 % der Einkommen untersucht, wobei nach Geschlecht und Herkunftsort unterschieden wurde. Die Ergebnisse sind in Abbildung 8 dargestellt.

Überraschend zeigt sich, dass sich die Streuung der Einkommensprämien für keine der vier untersuchten Gruppen zwischen 1991 und 2001 sichtbar zugenommen hat. Im Osten ist vor allem der starke Anstieg der Medianwerte interessant, welche die gestiegenen Einkommensprämien erklären. Zudem ist die Bewegung des untersten Quartils bei den Männern im Osten von einer negativen zu einer positiven Einkommensprämie ein wichtiges Signal für die gestiegene Wertschätzung für Hochschulbildung. Bei Männern und Frauen im Osten sinken die Einkommensprämien des obersten Quartils nach starkem Anstieg wieder ab, bleiben jedoch in beiden Fällen unter Westniveau.



Quelle: Mikrozensus

Im Westen Deutschlands hat die Streuung der Einkommensprämien ebenfalls nicht zugenommen. Für Männer und Frauen ist ein Absinken der Prämie des obersten Quartils zu beobachten, sodass folglich auch die gut verdienenden Akademikerinnen und Akademiker Einbußen zu beklagen hatten. Bei den Frauen ist die gleichbleibend hohe Differenz zwischen oberstem und unterstem Quartil zwischen 80 % und 90 % bemerkenswert.

Während im Westen die Einkommen im untersten Quartil im gesamten Zeitraum nur geringe oder negative Prämien zeigten, hat sich diese Entwicklung leider auch im Osten eingestellt. Das bedeutet, dass fast ein Viertel der Akademikerinnen und Akademiker weniger verdient als ein durchschnittlicher Absolvent/-inn einer Berufsausbildung. Für diese Hochschulabsolvent(inn)en hat sich folglich das Studium finanziell nicht ausgezahlt.

Insgesamt konnte somit zwar keine höhere Streuung der Einkommensprämien gezeigt werden, allerdings ist das allgemeine Absinken der Prämien im untersten Quartil bedenklich: Eine Einkommensprämie von unter 10 % für mehr als ein Viertel aller Akademikerinnen und Akademiker rechtfertigt weder auf persönlicher noch auf gesellschaftlicher Ebene die Sach- und Opportunitätskosten einer langwierigen Hochschulausbildung.

9 Zusammenfassung und Diskussion

Die möglichen Effekte einer Ausweitung der Bildungsbeteiligung hat Blossfeld als „Proletarisierung“, „Verdrängung“ und „Innovation“ beschrieben. Die oben beschriebenen Ergebnisse zeigen auf, welche dieser Szenarien eingetroffen sind.

Eine Proletarisierung ist glücklicherweise nicht eingetreten, da eine Hochschulausbildung immer noch ein guter Schutz vor Arbeitslosigkeit ist und sich dies auch durch eine Erhöhung der Akademikerquote nicht verändert hat. Arbeitslosigkeit trifft in der Regel nicht die Akademikerinnen und Akademiker, sondern die weniger Gebildeten.

Die Ausweitung der Bildungsbeteiligung in den 1990er Jahren hat im Westen Deutschlands zu einer Verdrängung weniger Qualifizierter aus ihren angestammten Tätigkeitsbereichen geführt. Die Einkommensprämien von Hochschulabsolvent(inn)en sind signifikant gefallen und für das untere Quartil der Akademikerinnen und Akademikern ist hinsichtlich der Einkommensprämie kein wesentlicher Vorteil gegenüber den Absolvent(inn)en einer beruflichen Bildung auszumachen.

Innovation führt oft dazu, dass der Bedarf an Akademikerinnen und Akademikern stärker ansteigt als die Nachfrage. Dieses Phänomen ließ sich im Osten Deutschlands in den 90er Jahren beobachten. Dort konnte die Nachfrage nach Akademikerinnen und Akademikern offenbar auch durch eine stark steigende Akademikerquote nicht ausgeglichen werden. Folglich stiegen die Einkommensprämien für Hochschulabsolvent(inn)en signifikant.

Zusammenfassend erhält Hochschulbildung in immer stärkerem Maß die Funktion eines Schutzes vor Arbeitslosigkeit, ein tatsächlicher bildungsadäquater Einkommensvorteil lässt sich für viele Absolvent(inn)en jedoch nicht mehr damit verbinden. Sinkende Einkommensprämien weisen zudem darauf hin, dass Hochschulabsolvent(inn)en zunehmend in Tätigkeitsfeldern arbeiten, für die sie eigentlich überqualifiziert sind.¹⁹⁾

Insofern könnte für viele Schüler zukünftig die Aufnahme einer nicht-akademischen Berufsausbildung wesentlich attraktiver als ein Studium werden. Benötigt der Arbeitsmarkt jedoch tatsächlich eine höhere Anzahl von Akademikerinnen und Akademikern, so sollte über höhere oder wenigstens konstante Einkommensprämien hier ein entsprechendes Signal an den Bildungsmarkt gegeben werden. Da die langwierige „Produktion“ von Akademikerinnen und Akademikern in Hochschulen zu einer Zeitverzögerung vor deren Eintritt ins Berufsleben führt, laufen die Unternehmen zunehmend noch mehr Gefahr, ihre Nachfrage nach qualifizierten Akademikerinnen und Akademikern nicht decken zu können.

¹⁹⁾ Vgl. Sichermann (1991), Dolton/Vignoles (1997).

Literatur

Acemoglu, D. (2003a): Patterns of skill premia. *Review of Economic Studies*, 70. Jg., S. 199 – 230.

Acemoglu, D. (2003b): Labor- and capital-augmenting technical change. *Journal of the European Economic Association* March, 1. Jg., Nr. 1, S. 1 – 37.

Ammermüller, A.; Dohmen, D. (2004): Individuelle und soziale Erträge von Bildungsinvestitionen. *Studien zum deutschen Innovationssystem 1 – 2004*. Forschungsinstitut für Bildungs- und Sozialökonomie. Köln, 2004.

Angrist, J. D.; Krueger, A. B. (1998): *Empirical strategies in labor economics*. Working Paper #401, Princeton University. Princeton, 1998.

Barro, R. J.; Sala-i-Martin, X. (1995): *Economic growth*. Cambridge, 1995.

Blien, U.; Reinberg, A.; Tessaring, M. (1990): Die Ermittlung der Übergänge zwischen Bildung und Beschäftigung. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 23. Jg., S. 181 – 204.

Blossfeld, P. (1983): Höherqualifizierung und Verdrängung, in: Haller, M.; Müller, W. (Hrsg.): *Beschäftigungssystem im gesellschaftlichen Wandel: Historische Entwicklungen und internationale Strukturunterschiede*. Frankfurt/New York, 1983.

Brauns, H.; Müller, W.; Steinmann, S. (1997): *Educational Expansion and Returns to Education – A Comparative Study on Germany, France, the UK, and Hungary*. Arbeitspapier Nr. 23, Mannheimer Zentrum für Europäische Sozialforschung. Mannheim, 1997.

Card, D. (2001): Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, 69. Jg., S. 1127 – 1160.

Card, D.; Lemieux, T. (2001) : Can falling supply explain the rising return to college for younger men? *The Quarterly Journal of Economics*, May 2001, S. 705 – 746.

Dolton, P. J.; Vignoles, A. (1997): Graduate overeducation: A European perspective. *Higher Education in Europe*. 22. Jg., Nr. 4, S. 475 – 484.

Fersterer, J.; Winter-Ebmer, R. (1999): Are Austrian returns to education falling over time? In: *Labour Economics*, 10. Jg., Nr. 1, S. 73 – 89.

Fitzenberger, B.; Garloff, A.; Kohn, K. (2003): *Beschäftigung und Lohnstrukturen nach Qualifikation und Altersgruppen: Eine empirische Analyse auf Basis der IAB-Beschäftigtenstichprobe*. Zentrum für empirische Wirtschaftsforschung, Discussion Paper 03 – 75. Mannheim, 2003.

Katz, L. F.; Murphy, K. M. (1992): Changes in relative wages 1963-1987: Supply and demand factors. *Quarterly Journal of Economics*, 57. Jg., S. 35 – 78.

Lui, H.-K.; Suen, W. (2004): The shrinking earnings premium for university graduates in Hong Kong: The effect of quantity or quality? *Contemporary Economic Policy*, 23. Jg., Nr. 2, S. 242 – 254.

Mincer, J. (1997): The production of human capital and the life cycle of earnings: Variations on a theme. *Journal of Labor Economics*, 15. Jg., Nr. 1, S. 26 – 47.

OECD (Hrsg., 1995): *Bildung auf einen Blick – OECD-Indikatoren 1995*. Deutsche Übersetzung durch das Bundesministerium für Bildung und Forschung. Organisation für wirtschaftliche Zusammenarbeit und Entwicklung. Paris, 1995.

OECD (Hrsg., 2004): *Bildung auf einen Blick – OECD-Indikatoren 2004*. Deutsche Übersetzung durch das Bundesministerium für Bildung und Forschung. Organisation für wirtschaftliche Zusammenarbeit und Entwicklung. Paris, 2004.

Reinberg, A; Fischer, G.; Tessaring, M. (1995): Auswirkungen der Bildungsexpansion auf die Erwerbs- und Nichterwerbstätigkeit. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 25. Jg., Nr. 3, S. 300 – 322.

Sichermann, N. (1991): Overeducation in the labor market. *Journal of Labor Economics*. 9. Jg., S. 101 – 122.

Steiner, V.; Lauer, C. (2000): *Private Erträge von Bildungsinvestitionen in Deutschland*. Zentrum für empirische Wirtschaftsforschung, Discussion Paper 00 – 18. Mannheim, 2000.

Autoren- und Kontaktangaben:

Dipl.-Kffr. Christiane Mück, Doktorandin am Arbeitsbereich Weiterbildung der Carl-von-Ossietzky-Universität Oldenburg; E-Mail: cmueck@whu.edu

Dipl.-Kffr. Karen Mühlenbein, Doktorandin am Arbeitsbereich Public Management der Universität Hamburg, Department für Wirtschaft und Politik; E-Mail: muehlenb@whu.edu

Möglichkeiten der Analyse von Arbeitsmarktübergängen mit Daten des Mikrozensus-Panels

Michael Konold

Der jährlich stattfindende Mikrozensus ist als rotierende Panelstichprobe angelegt. Die Daten einzelner Jahre wurden bislang aber nicht zu Längsschnitten zusammengeführt. Im Rahmen eines Kooperationsprojektes wird derzeit nun daran gearbeitet, dieses Potenzial zu erschließen. Aufbereitet wird konkret der Mikrozensus-Längsschnitt 1996 bis 1999. Es ergeben sich hierdurch neue Analysemöglichkeiten, weil sich mit diesen Daten z. B. auf Grund der sehr hohen Fallzahl Längsschnittuntersuchungen auch noch für relativ kleine Bevölkerungsgruppen durchführen lassen. Der vorliegende Beitrag berichtet über den Stand der Arbeiten und wird für den zentralen Bereich der Untersuchung von Arbeitsmarktübergängen das Analysepotenzial des Mikrozensus-Panels aufzeigen. Zwei Problemen widmet er dabei besondere Aufmerksamkeit: Das erste Problem besteht darin, dass Personen, die aus einem Auswahlbezirk fortziehen, nicht wiederbefragt werden und deshalb in Längsschnittanalysen nicht mit einbezogen werden können. Das zweite Problem ist der Sachverhalt, dass Verläufe im Mikrozensus-Panel in den meisten Fällen nicht durchgehend rekonstruierbar sind, weil Informationen über die Zeiten zwischen den einzelnen Mikrozensus-Berichtswochen in der Regel fehlen. Wie gezeigt werden wird, ergeben sich aus den beiden Punkten gewisse Einschränkungen. Das Analysepotenzial des Mikrozensus-Panels wird dadurch jedoch nicht grundsätzlich beeinträchtigt.

1 Einleitung

Zurzeit wird im Rahmen eines Kooperationsprojektes zwischen amtlicher Statistik und Wissenschaft die Basis dafür geschaffen, dass mit Daten des jährlichen Mikrozensus zukünftig auch Längsschnittanalysen durchgeführt werden können, wobei erste Daten bereits zusammengeführt wurden.¹⁾ Die Daten-Infrastruktur in Deutschland wird hierdurch verbessert, da Mikrozensus-Paneldaten Analysen ermöglichen, welche in dieser Form bisher nicht oder nur bedingt durchführbar waren. Drei „Stärken“ zeichnen die Längsschnitte aus:²⁾ Sie umfassen erstens sehr viele Personen, sodass Analysen auch noch für relativ kleine Bevölkerungsgruppen durchgeführt werden können. Zweitens ist die Zahl der Unit- und Item-Nonresponse-Fälle wegen der beim Mikrozensus gegebenen Auskunftspflicht im Vergleich zu freiwillig durchgeführten Erhebungen sehr gering. Drittens schließlich ist im Rahmen der Längsschnitte ein breites Merkmalspektrum verfügbar, welches vor allem im Bereich der Erwerbstätigkeit differenzierte Auswertungen erlaubt.

1) Das Projekt wird gemeinsam vom Statistischen Bundesamt, vom Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen, vom Zentrum für Umfragen, Methoden und Analysen (ZUMA) Mannheim sowie von der Freien Universität Berlin durchgeführt. Gefördert wird es vom Bundesministerium für Bildung und Forschung und von der Deutschen Forschungsgemeinschaft. – 2) Mit „Längsschnitt“ ist hier immer der Fall gemeint, dass Informationen von Personen aus zwei oder mehreren Jahren (Panel-Wellen) zu einem Datensatz zusammengeführt werden.

Den beschriebenen Vorteilen stehen zwei grundsätzliche Nachteile gegenüber: Der erste besteht darin, dass Personen, die aus einem Auswahlbezirk fortziehen, beim Mikrozensus nicht wiederbefragt werden. Stattdessen werden diejenigen Personen einbezogen, die in die ausgewählten Bezirke zuziehen (Prinzip der Flächenstichprobe).³⁾ Die Folge ist ein Prozess des Ausfalls räumlich mobiler Personen im Längsschnitt, der bei bestimmten Analysen zu Ergebnisverzerrungen führen kann (vgl. hierzu Stauder 2003, Zühlke 2003 sowie Basic et al. 2005). Der zweite Nachteil resultiert aus dem Berichtswochenkonzept des Mikrozensus. Dieses hat zur Folge, dass im Rahmen des Mikrozensus-Panels überwiegend Informationen verfügbar sind, die sich auf die jeweilige Mikrozensus-Berichtswoche beziehen. Informationen über Veränderungen in den Zeiträumen zwischen den einzelnen Berichtswochen liegen in der Regel nicht vor (vgl. hierzu Konold 2005a). Ein Problem ergibt sich hieraus, da Personen innerhalb eines Jahres in einem Bereich mehr als einen Übergang realisieren können. Zu denken wäre hier zum Beispiel an mehrfache Wechsel des Erwerbsstatus oder an mehrfache Veränderungen der Einkommenssituation. Im Rahmen des Mikrozensus-Panels werden also nicht alle Übergänge beobachtet. Von einigen wenigen Ausnahmen abgesehen ist zudem unbekannt, ab wann bzw. von wann bis wann einzelne Status eingenommen wurden.

Aus den beiden Punkten ergeben sich für Analysen mit Daten des Mikrozensus-Panels gewisse Einschränkungen, die im Folgenden für den zentralen Bereich der Untersuchung von Arbeitsmarktübergängen ausführlich diskutiert werden. Zunächst werden jedoch einige allgemeine Informationen zu den Daten und deren Verfügbarkeit gegeben. Informiert wird außerdem kurz über die wichtigsten Merkmale, die für Übergangsanalysen im Arbeitsmarktbereich zur Verfügung stehen.

2 Aufbereitung des Mikrozensus als Panelstichprobe

Die beim jährlich durchgeführten Mikrozensus gegebene sehr hohe Fallzahl,⁴⁾ die Auskunftspflicht sowie die Verfügbarkeit einer Vielzahl von Informationen aus Bereichen, in denen Längsschnittanalysen eine große Bedeutung zukommt, hat schon seit einiger Zeit den Wunsch nach einem Mikrozensus-Panel laut werden lassen. Möglich ist ein solches Panel, weil die beim Mikrozensus ausgewählten Bezirke vier Jahre lang in der Stichprobe verbleiben. Entsprechend werden alle Personen, die in den Auswahlbezirken leben, und bei denen die Möglichkeit hierzu besteht, viermal hintereinander befragt.⁵⁾ Es liegen also für die meisten Befragungspersonen Daten aus mehreren Jahren vor, welche miteinander verknüpft werden können.

Weil Auswahlbezirke immer vier Jahre im Mikrozensus verbleiben, können Mikrozensus-Längsschnitte für drei verschiedene Zeitspannen konstruiert werden: Möglich sind Zwei-Jahres-Längsschnitte, Drei-Jahres-Längsschnitte und Vier-Jahres-Längsschnitte, wobei diese jeweils unterschiedlich viele Personen umfassen. Dies hängt mit einem Rotationsverfahren zusammen, welches beim Mikrozensus zur Anwendung kommt (vgl. hierzu Abbildung 1).⁶⁾ Es sieht vor, dass je-

3) Beim Mikrozensus wird keine Stichprobe von Haushalten oder Personen gezogen, sondern es werden Bezirke ausgewählt, in denen alle dort wohnhaften Personen befragt werden. – 4) Befragt wird jedes Jahr etwa ein Prozent der Wohnbevölkerung Deutschlands. – 5) Nicht immer ist eine viermalige Befragung möglich. So kann eine Person die weitere Teilnahme verweigern, nicht angetroffen werden, sterben oder aus dem Auswahlbezirk fortziehen. Zu beachten ist weiterhin, dass eine Person erst im zweiten, dritten oder vierten Jahr zuziehen kann. – 6) Detaillierte Informationen hierzu finden sich in Zühlke (2001) sowie Heidenreich (2002).

des Jahr ein Viertel der Bezirke durch neue Bezirke ersetzt wird. Entsprechend kann ein balancierter Zwei-Jahres-Längsschnitt für deutlich mehr Personen erstellt werden, als ein balancierter Drei-Jahres-Längsschnitt (drei Rotationsviertel vs. zwei). Ein balancierter Drei-Jahres-Längsschnitt wiederum für deutlich mehr, als eine balancierter Vier-Jahres-Längsschnitt (zwei Rotationsviertel vs. eines). Selbst balancierte Vier-Jahres-Längsschnitte umfassen auf Bundesebene aber noch über 120 000 Personen.⁷⁾

Das zur Anwendung kommende Rotationsverfahren hat einen entscheidenden Vorteil: Es bewirkt, dass Längsschnitte für jeden beliebigen Zeitabschnitt konstruiert werden können. Würde nicht Jahr für Jahr ein Viertel der Bezirke ausgetauscht, sondern alle vier Jahre die Gesamtheit aller Bezirke durch eine neue Stichprobe ersetzt, so wäre dies nicht möglich.

Abb 1. Jahre, in denen die 1996 befragten Rotationsviertel in den Mikrozensus einbezogen wurden

Rotationsviertel (Kennung)	Jahre, in denen das Rotationsviertel in den Mikrozensus einbezogen wurde						
	1993	1994	1995	1996	1997	1998	1999
03 – 3	x	x	x	x			
03 – 4		x	x	x	x		
04 – 1			x	x	x	x	
04 – 2				x	x	x	x

Im Rahmen des eingangs erwähnten Kooperationsprojektes wird zurzeit der Vier-Jahres-Längsschnitt 1996 bis 1999 aufbereitet. Auf die Daten gerade dieser Jahre wird deshalb zurückgegriffen, weil der Längsschnitt 1996 bis 1999 den ersten konstruierbaren Vier-Jahres-Längsschnitt darstellt. Eine Zusammenführung älterer Daten ist nicht möglich, weil die Zuweisung einer identischen Ordnungsnummer an Haushalte über einzelne Jahre hinweg erst seit dem Mikrozensusgesetz vom 16. Januar 1996 zulässig ist und folglich erst ab 1996 umgesetzt wurde.

Der Längsschnitt 1996 bis 1999 wird Anfang 2006 als Scientific-Use-File (SUF) verfügbar sein. Schwächer anonymisierte Daten, die zusätzliche Auswertungsmöglichkeiten bieten, können am Gastwissenschaftlerarbeitsplatz in einem Standort des Forschungsdatenzentrums der Statistischen Landesämter oder im Forschungsdatenzentrum des Statistischen Bundesamtes ausgewertet werden. Möglich ist zudem eine Nutzung mittels kontrollierter Datenfernverarbeitung. Mit der Aufbereitung eines weiteren Längsschnitts (2001 bis 2004) wird aller Voraussicht nach im Jahr 2006 begonnen.

3 Analyse von Arbeitsmarktübergängen – verfügbare Merkmale

Da Fragen zur Erwerbstätigkeit im Mikrozensus besonders breiten Raum einnehmen und das Spektrum an Variablen, die sich im Zusammenhang mit Arbeitsmarktprozessen als erklärende Variablen anbieten groß ist, bestehen im Mikrozensus-Panel besonders vielfältige Möglichkeiten der Analyse von Arbeitsmarktübergängen. Im Folgenden wird über die wichtigsten der verfügbaren Merkmale ein kurzer Überblick gegeben, um einen Eindruck über das bestehende Potenzial des Mikrozensus-Panels in diesem Bereich zu vermitteln.⁸⁾

7) Dies gilt für nicht anonymisierte Datensätze. Für Scientific-Use-Files wird sich voraussichtlich ein Umfang von ungefähr 90 000 Personen ergeben. – 8) Die Angaben beziehen sich auf die Jahre 1996 bis 1999. In späteren Jahren gibt es einige Abweichungen im Detail. Im Wesentlichen gelten die Aussagen jedoch auch für spätere Jahre.

Verfügbar ist im Mikrozensus-Panel zunächst einmal für jedes Jahr der Erwerbsstatus, die Erwerbsbeteiligung und die Stellung im Beruf. Was den Erwerbsstatus angeht, so kann zwischen Erwerbstätigen, Erwerbslosen und Nichterwerbspersonen gemäß der Definition der International Labor Organization (ILO) unterschieden werden (vgl. hierzu Schmidt 2000). Es bestehen auf Grund des Vorliegens zahlreicher Einzelinformationen außerdem verschiedene Möglichkeiten, die genannten Erwerbsstatus in anderer Weise voneinander abzugrenzen.⁹⁾

Die Erwerbsbeteiligung wird sehr differenziert erfasst. So lässt sich ermitteln, ob eine Person in einer Berichtswoche Vollzeit erwerbstätig, Teilzeit erwerbstätig, geringfügig beschäftigt oder überhaupt nicht erwerbstätig war. Darüber hinaus steht in allen Jahren die Information zur Verfügung, wie viele Tage und Stunden eine Person in der Woche normalerweise gearbeitet hat und wie viele Arbeitstage und Arbeitsstunden von ihr in der jeweiligen Berichtswoche geleistet wurden.

Zur Stellung im Beruf liegen für jedes Erhebungsjahr ebenfalls genaue Informationen vor. Unterschieden wird hier zwischen Selbständigen mit Beschäftigten, Selbständigen ohne Beschäftigte, mithelfenden Familienangehörigen, Beamten, Angestellten, Arbeitern, kaufmännisch-technischen Auszubildenden, gewerblichen Auszubildenden, Soldaten, Grundwehrdienstleistenden und Zivildienstleistenden.

Weitere Möglichkeiten für Übergangsanalysen ergeben sich aus folgenden Mikrozensus-Merkmalen:

- Wirtschaftszweig der ausgeübten Tätigkeit gemäß dem systematischen Verzeichnis der Wirtschaftszweige des Statistischen Bundesamtes
- Beruf gemäß der Klassifikation der Berufe des Statistischen Bundesamtes
- Berufswechsel im letzten Jahr
- Betriebs-/Firmenwechsel im letzten Jahr
- Zeitpunkt der Aufnahme der Tätigkeit beim derzeitigen Arbeitgeber/als Selbständiger (Jahr, Monat)
- Beendigung der letzten bzw. der früheren Erwerbstätigkeit (Jahr, Monat)¹⁰⁾
- Dauer der Arbeitssuche¹¹⁾
- Situation vor der Arbeitssuche
- Art des Arbeitsvertrages

Die Liste zeigt, dass mit Daten des Mikrozensus-Panels z. B. auch Arbeitgeberwechsel untersucht werden können. Sie zeigt weiterhin, dass das Mikrozensus-Panel zwar nur wenige Verlaufs- bzw. Retrospektivinformationen bereitstellt, dass in einigen Fällen solche Informationen jedoch vorliegen und Analysemöglichkeiten eröffnen.¹²⁾

9) Nicht vorgenommen werden kann eine Abgrenzung der Arbeitslosen gemäß der Definition der Bundesagentur für Arbeit (vormals: Bundesanstalt für Arbeit). – 10) Die Frage nach der Beendigung der letzten bzw. der früheren Erwerbstätigkeit wird leider nur Personen gestellt, die in der Berichtswoche nicht erwerbstätig waren. Sie wird außerdem nur an eine Unterstichprobe gerichtet. – 11) Vorgegeben sind bei dieser Frage acht Antwortkategorien. Die Dauer liegt also nur klassiert vor. – 12) Auf Variablen, die sich als erklärende Variablen in multivariaten Modellen anbieten, soll hier nicht im Detail eingegangen werden. Hingewiesen sei lediglich darauf, dass neben sozio-demographischen und sozio-ökonomischen Basismerkmalen u. a. Informationen zur Arbeitssuche vorliegen.

4 Probleme bei der Analyse von Arbeitsmarktübergängen

4.1 Ausfall räumlich mobiler Personen im Längsschnitt

Personen, die aus einem Auswahlbezirk fortziehen, werden beim Mikrozensus nicht erneut befragt. Daher können in Längsschnittanalysen auf Basis des Mikrozensus nur Personen einbezogen werden, die im Betrachtungszeitraum nicht umgezogen sind. Die Beschränkung der Untersuchungsgruppe auf räumlich immobile Personen kann zu Ergebnisverzerrungen führen, sofern die räumliche Mobilität mit dem Untersuchungsmerkmal zusammenhängt.¹³⁾ Bereits vorliegende Studien weisen zum Beispiel darauf hin, dass zwischen Übergängen in Erwerbstätigkeit und räumlicher Mobilität ein positiver Zusammenhang besteht (vgl. Zühlke 2003 sowie Basic et al. 2005). Wird der Prozentsatz der Personen mit Übergang in Erwerbstätigkeit lediglich auf der Basis räumlich immobiler Personen bestimmt, so kann dies folglich zu einer Unterschätzung des wahren Prozentwertes führen.

Multivariate Analysen sind von der Problematik ebenfalls betroffen. Die Voraussetzungen für Verzerrungen sind hier jedoch komplexer. An einem Beispiel soll dies veranschaulicht werden: Betrachtet wird der Fall einer Analyse von Übergängen aus Erwerbslosigkeit in Erwerbstätigkeit, bei der das Geschlecht als erklärende Variable genutzt wird. Der für das Geschlecht geschätzte Effekt kann hier unter der Voraussetzung verzerrt werden, dass Männer und Frauen sich im Hinblick auf den statistischen Zusammenhang, der zwischen räumlicher Mobilität und Übergängen aus Erwerbslosigkeit in Erwerbstätigkeit besteht, unterscheiden. Dies wäre zum Beispiel dann der Fall, wenn Männer, die aus Erwerbslosigkeit in Erwerbstätigkeit übergehen, häufiger räumlich mobil sind als Frauen und gleichzeitig zwischen den erwerbslosen Männern und Frauen, die nicht in Erwerbstätigkeit wechseln, kein Unterschied bei der Umzugswahrscheinlichkeit besteht. Für Verzerrungen müssen also relativ spezielle Voraussetzungen vorliegen. Eine nahe liegende Vermutung ist deshalb, dass sich der Ausfall räumlich mobiler Personen bei multivariaten Analysen in deutlich geringerem Maße auswirkt als bei der Bestimmung von Übergangswahrscheinlichkeiten.

Im Folgenden sollen zu der Problematik des Ausfalls räumlich mobiler Personen einige empirische Ergebnisse präsentiert werden. Betrachtet werden zunächst die Auswirkungen auf Übergangswahrscheinlichkeiten. Danach wird in einem zweiten Schritt der Blick auf multivariate Analysen gerichtet.

4.1.1 Auswirkungen auf geschätzte Übergangswahrscheinlichkeiten

Tabelle 1 zeigt für einige ausgewählte Arbeitsmarktübergänge auf, in welchem Maße der Prozentsatz der 18- bis 65-jährigen Personen mit Übergang unter- bzw. überschätzt wird, wenn eine Analyse lediglich mit räumlich immobilen Personen durchgeführt wird. Sie informiert ferner darüber, in welchem Maße Verzerrungen zunehmen, wenn das betrachtete Zeitintervall von einem auf zwei bzw. drei Jahre ausgeweitet wird.¹⁴⁾ Für die Auswertung wurden Daten des Sozio-

13) In Deutschland ist in den Jahren 1996, 1997, 1998 und 1999 jeweils knapp ein Zehntel aller Personen umgezogen. Die Zahl der räumlich mobilen Personen ist also nicht vernachlässigbar. – 14) Die ausgewiesenen Ergebnisse beziehen sich auf Analysen, bei denen der Status von Personen im Jahr 1996 und der Status von Personen im Jahr 1997, 1998 bzw. 1999 betrachtet wurde. Andere Analysevarianten erbringen von einigen wenigen Ausnahmen abgesehen ähnliche Ergebnisse.

oekonomischen Panels (SOEP) herangezogen, da das Mikrozensus-Panel selbst nur sehr begrenzte Möglichkeiten für eine solche Analyse bietet.

Das SOEP eignet sich für die Untersuchung der Verzerrungen aus drei Gründen: Erstens stellt das SOEP eine repräsentative Studie dar und bietet ausreichende Fallzahlen, um auch noch solche Übergänge abbilden zu können, die von relativ wenigen Personen vollzogen werden. Zweitens werden beim SOEP Personen, die umziehen, erneut befragt. Es kann also für alle Übergänge im SOEP untersucht werden, welche Folgen es hat, wenn räumlich mobile Personen nicht in eine Analyse einbezogen werden.¹⁵⁾ Drittens schließlich bietet das SOEP ein ähnliches Merkmalspektrum wie das Mikrozensus-Panel. Mit dem SOEP lassen sich aus diesem Grund zahlreiche Analysen nachbilden, die mit dem Mikrozensus-Panel durchgeführt werden könnten.

Die Spalte 1 der Tabelle 1 zeigt, wie viel Prozent der betrachteten Personen den betreffenden Übergang gemäß SOEP jeweils vollzogen haben. Die Spalte 2 weist aus, welcher Wert sich ergibt, wenn eine Analyse nur mit Personen durchgeführt wird, die im Betrachtungszeitraum nicht umgezogen sind. Der Spalte 3 kann entnommen werden, wie groß die Über-/Unterschätzung des für alle Personen geschätzten Wertes in Prozentpunkten ausfällt, wenn in eine Analyse nur Personen ohne Umzug einbezogen werden (absoluter Fehler). Aus der Spalte 4 geht hervor, welchen Wert die Über-/Unterschätzung in Prozent annimmt (relativer Fehler).

Die Spalten 3 und 4 zeigen, dass bei Arbeitgeberwechseln und Übergängen aus Erwerbslosigkeit in Erwerbstätigkeit ein relativ großer Fehler realisiert wird, wenn die Bestimmung des Prozentsatzes der Personen mit Übergang auf Basis räumlich immobiler Personen erfolgt. So wird der Prozentsatz der Personen, die 1997 bei einem anderen Arbeitgeber beschäftigt waren als im Jahr 1996 gemäß SOEP z. B. um 13,1 Prozent unterschätzt, wenn in eine Analyse nur räumlich immobile Personen einbezogen werden. Die beiden Spalten zeigen weiterhin, dass die Verzerrungen bei diesen beiden Übergängen bei Ausweitung des betrachteten Zeitintervalls deutlich zunehmen. Beim Übergang „Arbeitgeberwechsel“ erhöht sich die Unterschätzung in Prozent von 13,1 (1996/97) auf 25,0 (1996/99). Bei Übergängen aus Erwerbslosigkeit in Erwerbstätigkeit steigt sie von 7,7 (1996/97) auf 19,0 (1996/99) an. Die Ergebnisse sind inhaltlich plausibel, da einiges für die Annahme spricht, dass ein Wechsel des Arbeitgebers ebenso wie ein Übergang aus Erwerbslosigkeit in Erwerbstätigkeit relativ häufig mit einem Umzug verbunden ist.¹⁶⁾ Ähnliche Werte ergeben sich für den Übergang „Nichterwerbsperson → erwerbstätige Person“, der aus Platzgründen nicht in die Tabelle aufgenommen wurde.

15) Dem Sachverhalt, dass räumlich mobile Personen eine erhöhte Panel-Ausfallwahrscheinlichkeit besitzen, wird im SOEP dadurch begegnet, dass das Merkmal Mobilität bei der Berechnung der Längsschnittgewichte Berücksichtigung findet (vgl. hierzu Frick 1996, S. 127ff). – 16) Um zu prüfen, ob die betrachteten Zeiträume möglicherweise einen Ausnahmefall darstellen, wurden entsprechende Analysen auch mit gepoolten Längsschnitten durchgeführt (Gepoolt wurden z. B. die drei Ein-Jahres-Längsschnitte 1996 – 97, 1997 – 98 und 1998 – 99). Für die betrachteten Übergänge erbrachten diese Analysen durchgehend vergleichbare Resultate. Ein Hausman-Test, der allerdings nur ungewichtet, also ohne Gewichtung mit den im SOEP bereitgestellten Faktoren durchführbar ist, weist die Differenz zwischen der für alle Personen und der für die immobilen Personen geschätzten Übergangswahrscheinlichkeit bei den ersten beiden Übergängen in der Tabelle für alle Zeitintervalle als signifikant aus (Zum Hausman-Test vgl. Basic et al. 2005, S. 173ff).

1. Fehler, der bei der Bestimmung des Prozentsatzes der Personen mit Übergang realisiert wird, wenn in eine Analyse lediglich räumlich immobile Personen einbezogen werden: Übersicht für ausgewählte Übergänge aus dem Bereich der Erwerbstätigkeit¹⁾

Betrachteter Übergang	Personen ²⁾ unter Risiko ³⁾ , die den betreffenden Übergang realisiert haben, in Prozent Geschätzter Wert auf Basis ...		Wert Sp. 2 minus Wert Sp. 1	Wert Sp. 3 dividiert durch Wert Sp. 1 * 100
	aller Personen	immobiler Personen		
	1	2	3	4
Arbeitgeberwechsel				
1996 1997	6,8	5,9	-0,9	-13,1
1996 1998	12,2	10,3	-1,9	-15,6
1996 1999	17,8	13,4	-4,5	-25,0
Erwerbslos → Erwerbstätig				
1996 1997	32,6	30,1	-2,5	-7,7
1996 1998	37,1	29,8	-7,3	-19,6
1996 1999	39,6	32,0	-7,5	-19,0
Erwerbslos → Erwerbslos				
1996 1997	51,8	52,8	1,0	2,0
1996 1998	42,0	46,5	4,5	10,7
1996 1999	31,3	33,0	1,7	5,3
Erwerbstätig → Erwerbslos				
1996 1997	5,2	5,4	0,2	2,9
1996 1998	6,6	6,0	-0,6	-9,0
1996 1999	5,8	6,3	0,4	7,2
Vollzeit → Teilzeit				
1996 1997	1,4	1,4	0,0	1,4
1996 1998	2,3	2,3	0,0	-0,7
1996 1999	2,8	2,9	0,0	1,6

1) Betrachtet wurde jeweils der Status, den eine Person im Ausgangsjahr (1996) innehatte, und der Status, der von ihr im zweiten betrachteten Jahr (1997, 1998 bzw. 1999) eingenommen wurde. Die Werte in den Spalten 3 und 4 wurden auf Basis der ungerundeten Werte berechnet. – 2) Personen, die im Jahr 1996 im Alter zwischen 18 und 65 Jahren waren und welche im Zeitraum 1996 bis 1997, 1996 bis 1998 bzw. 1996 bis 1999 in einem Privathaushalt in Deutschland gelebt haben. – 3) Personen, die im Jahr 1996 den betreffenden Ausgangsstatus innehatten und damit dem Risiko „ausgesetzt“ waren, den betreffenden Übergang zu vollziehen. – – – Daten: Das Sozio-oekonomische Panel (SOEP) 1996 bis 1999; eigene Berechnungen; hochgerechnete Ergebnisse

Für die weiteren in Tabelle 1 aufgeführten Übergänge ergibt sich ein anderes Bild: Bei Übergängen aus Erwerbstätigkeit in Erwerbslosigkeit und im Fall eines Verbleibs in Erwerbslosigkeit treten vergleichsweise geringe Verzerrungen auf. Bei der Bestimmung des Prozentsatzes der Personen, die einen Übergang aus Vollzeit- in Teilzeit-Erwerbstätigkeit vollzogen haben, wirkt sich eine Beschränkung der Untersuchungsgruppe auf räumlich immobile Personen gemäß SOEP praktisch gar nicht aus.

Da jedoch zwischen einigen Arbeitsmarktübergängen und räumlicher Mobilität Zusammenhänge bestehen, stellt sich die Frage, ob durch eine Gewichtung Ausfälle, die aus Umzügen resultieren, kompensiert werden können. Im Rahmen des bereits erwähnten Kooperationsprojektes „Aufbereitung und Bereitstellung des Mikrozensus als Panel-Stichprobe“ wurden Untersuchungen hierzu durchgeführt. Die vom Statistischen Bundesamt und der Freien Universität Berlin entwickelten Gewichte haben sich dabei in Tests als wirkungsvoll erwiesen (vgl. Konold 2005b). Sie erlauben zwar keine vollständige Kompensation der Ausfälle. In den meisten Fällen sind sie jedoch in der Lage, den Ausfall der räumlich mobilen Personen in hohem Maße zu kompensieren. Eine Veröffentlichung zu diesem Komplex, in der neben zahlreichen Arbeitsmarktübergängen auch Übergänge aus dem Bereich Partnerschaft/Familie betrachtet werden, ist in Vorbereitung.

4.1.2 Auswirkungen auf die Ergebnisse multivariater Analysen

Multivariate Logit-Modelle zur Analyse von Übergängen aus Erwerbslosigkeit in Erwerbstätigkeit zeigen, dass sich eine Nichtberücksichtigung räumlich mobiler Personen dort nur in geringem Maße auswirkt (vgl. Basic et al. 2005 sowie Möhring und Zühlke 2005). In diesem Abschnitt wird ergänzend hierzu ein Logit-Modell zur Analyse von Übergängen aus abhängiger Beschäftigung in Erwerbslosigkeit betrachtet. Die Untersuchungen wurden ebenfalls mit Daten des SOEP durchgeführt.

Tabelle 2 weist für das betrachtete Modell die Ergebnisse für drei unterschiedlich große Zeitintervalle aus. Ein fett gedrucktes Wertepaar zeigt einen Unterschied zwischen dem geschätzten Koeffizienten bei der Analyse mit allen Personen und dem geschätzten Koeffizienten bei der Analyse mit den Immobilen hinsichtlich des Vorzeichens und/oder der Signifikanz an.¹⁷⁾ Ein grauer Hintergrund weist darauf hin, dass gemäß Hausman-Test zwischen den beiden geschätzten Koeffizienten eine signifikante Differenz besteht.¹⁸⁾

Tabelle 2 belegt, dass sich die Ergebnisse einer Analyse mit allen Personen nur in geringem Maße von den Ergebnissen einer Analyse auf Basis immobiler Personen unterscheiden. So tritt z. B. bei keiner einzigen Variable ein unterschiedliches Vorzeichen auf. In den Fällen, in denen sich ein geschätzter Koeffizient bei der Analyse mit allen Personen als signifikant erweist, bei der Analyse mit den immobilen Personen hingegen nicht (oder umgekehrt), liegen die geschätz-

2. Determinanten der Wahrscheinlichkeit eines Übergangs aus abhängiger Beschäftigung in Erwerbslosigkeit: Gegenüberstellung der Ergebnisse logistischer Regressionsanalysen, die mit allen Personen durchgeführt wurden und der Ergebnisse logistischer Regressionsanalysen, bei denen lediglich räumlich immobile Personen betrachtet wurden ¹⁾						
Unabhängige Variable ²⁾	Geschätzter Wert des Koeffizienten (β) ³⁾					
	1996 → 1997 Analyse mit		1996 → 1998 Analyse mit		1996 → 1999 Analyse mit	
	allen Personen	immobilen Personen	allen Personen	immobilen Personen	allen Personen	immobilen Personen
	1	2	3	4	5	6
Konstante	-0,72**	-0,65*	-0,71**	-0,87**	-0,84**	-1,00**
Geschlecht: männlich	0,13	0,15	0,23*	0,36*	0,00	0,00
Staatsangehörigkeit: deutsch	-0,34	-0,39*	-0,36*	-0,33	-0,35	-0,40
Alter: 18 bis 30	-0,31	-0,38*	-0,15	-0,07	-0,27	-0,01
Alter: 51 bis 65	0,87***	0,83***	0,98***	0,99**	1,14***	1,18***
Beruflicher Bildungsabschluss (Fach-)Hochschulabschluss	-0,49**	-0,45*	-0,48**	-0,56**	-0,53**	-0,54**
Im öffentlichen Dienst	-1,39***	-1,34***	-1,02***	-0,98**	-1,21***	-1,03***
Betriebsgröße: 20 oder mehr	-0,11	-0,09	-0,21	-0,16	-0,36*	-0,48**
Monate beim Arbeitgeber > 36	-0,25	-0,34*	-0,21	-0,25	-0,18	-0,09
Wohnsitz in Westdeutschland	-1,28***	-1,25***	-0,97***	-0,82***	-0,53***	-0,44**
	-0,92***	-0,93***	-1,02***	-1,12***	-1,04***	-1,02***
n	5 346	4 831	4 959	4 012	4 675	3 476
Nagelkerkes R ²	0,12	0,12	0,11	0,11	0,10	0,10

1) Personen, die im Ausgangsjahr im Alter zwischen 18 und 65 Jahren waren und welche 1996 – 1997, 1996 – 1998 bzw. 1996 – 1999 in einem Privathaushalt in Deutschland gelebt haben. Betrachtet wurde jeweils der Status von Personen zum Zeitpunkt der Erhebung im Jahr 1996 und zum Zeitpunkt der Erhebung im Jahr 1997, 1998 bzw. 1999. Beamte und Wehr- und Zivildienstleistende wurden aus der Analyse ausgeschlossen. – 2) Bei allen Variablen handelt es sich um 0/1-kodierte Dummy-Variablen (1 entspricht Ja). Alter: Ref. = 31 bis 50; Abschluss: Ref. = kein Berufsabschluss. Bezugspunkt ist jeweils das Ausgangsjahr. – 3) ***: Koeffizient signifikant auf 0,1%-Niveau; **: ~ signifikant auf 1%-Niveau; *: ~ signifikant auf 5%-Niveau. – – – Daten: Das Sozio-oekonomische Panel (SOEP) 1996 bis 1999; eigene Berechnungen

17) Bei der Signifikanz wurde auf den Unterschied „signifikant/nicht signifikant“ abgestellt. – 18) Vgl. zu diesem Test Basic et al. (2005), S. 173ff. Ivo Marek danke ich für die Bereitstellung eines SPSS-Makros zur Berechnung der Test-Werte.

ten Werte relativ nah beieinander. Gemäß den Zahlen aus der Tabelle ist es also unwahrscheinlich, dass bei einer Analyse auf Basis immobiler Personen falsche Schlussfolgerungen gezogen würden. Der durchgeführte Hausman-Test weist in insgesamt drei Fällen eine Differenz als signifikant aus. Die absolute Differenz zwischen den geschätzten Koeffizienten nimmt allerdings nur in einem Fall einen größeren Betrag an, und zwar bei der Variable „Alter: 18 bis 30“ bei Betrachtung des Drei-Jahres-Intervalls 1996/99.

Auf ein Problem muss an dieser Stelle aufmerksam gemacht werden: Die Analyse, deren Ergebnisse in Tabelle 2 dargestellt sind, wurde ungewichtet durchgeführt, weil bei einer Gewichtung mit den im SOEP bereitgestellten Faktoren die ermittelten Standardfehler und Signifikanzen nur mit großer Vorsicht interpretiert werden können.¹⁹⁾ Ein Problem ergibt sich hieraus insofern, als räumlich mobile Personen eine erhöhte Panel-Ausfallwahrscheinlichkeit besitzen und deshalb im SOEP (leicht) unterrepräsentiert sind.²⁰⁾ Wird die Logit-Analyse gewichtet durchgeführt, so nehmen die Differenzen zwischen den geschätzten Koeffizientenwerten etwas zu. Es ergibt sich jedoch kein grundlegend anderes Bild. Hinzuweisen wäre ferner darauf, dass bei einer Analyse, bei der die Längsschnitte 1996 – 97, 1997 – 98 und 1998 – 99 gepoolt werden, die Differenzen geringer ausfallen als bei der Analyse für das Zeitintervall 1996/97. Die in Tabelle 2 sichtbaren Unterschiede zwischen den beiden Analysen scheinen also zumindest zum Teil den nicht allzu großen Fallzahlen geschuldet zu sein. Hierfür spricht auch, dass Differenzen eher unsystematisch bei verschiedenen unabhängigen Variablen auftreten.

4.2 Folgen des Berichtswochenkonzeptes

Bei Längsschnittanalysen stehen häufig direkte Übergänge oder der Verbleib in einem bestimmten Status im Blickpunkt. Das Interesse gilt also beispielsweise direkten Übergängen aus Erwerbslosigkeit in Erwerbstätigkeit oder der Frage, wie viel Prozent aller im Ausgangsjahr erwerbslosen Personen ein Jahr später immer noch erwerbslos waren. Betrachtet werden bei Längsschnittanalysen außerdem häufig „Ankunftszeiten“ bzw. Verweildauern. Ein Beispiel für eine solche Analyse wäre die Klärung der Frage, inwiefern sich verschiedene Vergleichsgruppen im Hinblick auf die durchschnittliche Zeitdauer unterscheiden, die bis zu einem Übergang in Erwerbstätigkeit vergeht.

Im Gegensatz zu Analysen, bei denen der Status in den Blick genommen wird, den Personen zu bestimmten Erhebungszeitpunkten inne hatten, gilt für diese beiden Arten von Längsschnittanalysen, dass die Auswertungsmöglichkeiten mit dem Mikrozensus-Panel beschränkt sind. So gibt es zunächst eine Problematik unbeobachteter Übergänge. Bei der Betrachtung von Verweildauern besteht zusätzlich das Problem, dass diese im Mikrozensus-Panel nur in einigen speziellen Fällen bestimmt werden können, und dass eine solche Bestimmung in diesen Fällen außerdem auch nur näherungsweise möglich ist.

Was mit unbeobachteten Übergängen gemeint ist, soll mit Hilfe eines einfachen Beispiels veranschaulicht werden:

19) Dies gilt auch für den Fall, dass die Gewichtungsfaktoren standardisiert werden. – 20) Der Sachverhalt, dass Ausländer und in Ostdeutschland lebende Personen im SOEP überrepräsentiert sind, stellt deshalb kein Problem dar, weil Staatsangehörigkeit und Wohnort als Kontrollvariablen in das Modell einbezogen werden können.

Angenommen, eine Person war gemäß Mikrozensus zunächst erwerbstätig. Im Folgejahr hatte sie dann den Status „Nichterwerbsperson“ inne. In diesem Fall wäre nicht sicher, dass die betreffende Person einen *direkten* Übergang aus Erwerbstätigkeit in den Zustand „Nichterwerbsperson“ vollzogen hat. Der Erwerbsstatus kann innerhalb eines Jahres mehrfach gewechselt werden. Folglich ist also z. B. denkbar, dass die betreffende Person erst eine Zeit lang erwerbslos war, bevor sie sich vom Arbeitsmarkt zurückgezogen hat.

Unbeobachtete Übergänge stellen deshalb ein Problem dar, weil bei Untersuchungen, die direkte Übergänge oder den Verbleib in einem bestimmten Status in den Blick nehmen, Ergebnisverzerrungen resultieren können. Genauer gesagt, sie zwingen den Forscher/die Forscherin mit bestimmten Annahmen zu operieren. Im Normalfall wäre dies – bezogen auf das Mikrozensus-Panel – die Annahme, dass Personen zwischen zwei Mikrozensus-Berichtswochen maximal einen Übergang realisieren. Bei einigen Fragestellungen müsste bei einer Analyse mit Daten des Mikrozensus-Panels eine solche Annahme getroffen werden.

Im Folgenden werden einige wichtige Ergebnisse durchgeführter Untersuchungen zu unbeobachteten Übergängen im Mikrozensus-Panel kurz zusammengefasst (vgl. ausführlicher hierzu Konold 2005a). Danach wird erörtert, welche Verweildauern im Mikrozensus-Panel betrachtet werden können.

4.2.1 Unbeobachtete Übergänge

Wie bei der Problematik des Ausfalls räumlich mobiler Personen im Längsschnitt, so gilt auch für die Problematik unbeobachteter Übergänge, dass eine Untersuchung der Folgen mit dem Mikrozensus-Panel selbst nicht möglich ist. Es muss auf andere Paneldaten zurückgegriffen werden, die es ermöglichen, die entsprechenden Werte für das Mikrozensus-Panel zu schätzen. Untersuchungen, die zu diesem Komplex mit Daten des SOEP durchgeführt wurden, zeigen, dass die Annahme, dass Personen innerhalb eines Jahres maximal einen Übergang realisieren, beim Familienstand in aller Regel zutrifft. Im Erwerbsbereich trifft diese Annahme jedoch häufig nicht zu. Als Hauptproblem erweist sich der Sachverhalt, dass für Personen, die zu zwei aufeinander folgenden Mikrozensus-Erhebungen jeweils den Status „erwerbstätig“ oder den Status „Nichterwerbsperson“ inne hatten, nicht ermittelt werden kann, ob sie zwischenzeitlich einmal erwerbslos gewesen sind. Es besteht lediglich die Möglichkeit, für einen Teil dieser Personen eine Betroffenheit von Erwerbslosigkeit auszuschließen. Analysen mit Daten des SOEP zeigen, dass von den Personen im Alter zwischen 18 und 65 Jahren 2,8 Prozent eine Episode der Arbeitslosigkeit²¹⁾ aufweisen, die zwischen April 1996 und April 1997 beginnt und endet. Für den Zeitraum April 1997 bis April 1998 ergibt sich ein Prozentwert von 2,9, für den Zeitraum April 1998 bis April 1999 ein Prozentwert von 2,4 (vgl. hierzu Tabelle 3).²²⁾ Folglich kann ein substanzieller Fehler auftreten, wenn bei einer Analyse von Übergängen in Erwerbslosigkeit mit Daten des Mikrozensus-Panels unterstellt wird, dass unbeobachtete Episoden der Erwerbslosigkeit

21) Die Verlaufsdaten des SOEP informieren darüber, von wann bis wann eine Befragungsperson arbeitslos gemeldet war. Eine Übertragung der Ergebnisse auf den Fall der Erwerbslosigkeit ist nicht völlig problematisch, aber zulässig. Insgesamt wird die Zahl unbeobachteter Erwerbslosigkeits-Episoden auf Basis des SOEP hochwahrscheinlich etwas unterschätzt (vgl. hierzu Jürges 2005). – 22) Der April wurde deshalb als Bezugsmonat gewählt, weil die Mikrozensus-Berichtswoche in den Jahren 1996 bis 1999 immer im April lag.

nicht vorkommen. Hierzu sei ein Beispiel gegeben: Werden Personen betrachtet, die im April 1996 erwerbstätig waren und wird gefragt, wie viel Prozent dieser Personen innerhalb von einem Jahr einen Übergang in Erwerbslosigkeit vollzogen haben, so wird der wahre Prozentwert gemäß SOEP um 2,8 Prozentpunkte bzw. 38,2 Prozent unterschätzt. Für den Drei-Jahres-Zeitraum 1996 bis 1999 ergibt sich eine Unterschätzung um 5,2 Prozentpunkte bzw. 34,4 Prozent (vgl. Tabelle 4).

Weil bei multivariaten Analysen in der Regel Unterschiede zwischen Vergleichsgruppen im Blickpunkt stehen, spricht einiges für die Annahme, dass diese von der Problematik unbeobachteter Übergänge in deutlich geringerem Maße betroffen sind. Durchgeführte Analysen mit Daten des SOEP haben allerdings gezeigt, dass sich Probleme aus dem Sachverhalt ergeben können, dass Männer deutlich häufiger unbeobachtete Erwerbslosigkeitsepisoden aufweisen als Frauen (vgl. Konold 2005a).

3. Personen mit Erwerbsstatus-Episoden, die zwischen zwei aufeinander folgenden Aprilmonaten beginnen und enden (in Prozent)			
Betrachteter Zeitraum	Personen ¹⁾ mit zumindest einer ...		
	Erwerbstätigkeits-Episode	Arbeitslosigkeits-Episode	Episode als Nichterwerbsperson
	1	2	3
1996 – 1997	1,8	2,8	1,2
1997 – 1998	1,5	2,9	1,2
1998 – 1999	2,1	2,4	1,2

1) Personen, die im Zeitraum 1996 bis 1998, 1997 bis 1999 bzw. 1998 bis 2000 in einem Privathaushalt in Deutschland gelebt haben und 1996, 1997 bzw. 1998 im Alter zwischen 18 und 65 Jahren waren – – – Datenquelle: SOEP (Episodendaten); eigene Berechnungen; hochgerechnete Ergebnisse

4. Prozent Erwerbstätige, die im Zeitraum von einem, zwei bzw. drei Jahren einen Übergang in Arbeitslosigkeit vollzogen haben nach unterschiedlichen Erhebungs- bzw. Analysekonzepten¹⁾				
Betrachteter Zeitraum	Erwerbstätige mit Übergang in Arbeitslosigkeit in Prozent gemäß		Wert Sp. 2 minus Wert Sp. 1	Wert Sp. 3 dividiert durch Wert Sp. 1 * 100
	Analyse auf Basis monatlicher Statusinformationen ²⁾	Analyse mit Statusinformationen im Jahresabstand ³⁾		
	1	2	3	4
1996 – 1997	7,4	4,6	-2,8	-38,2
1996 – 1998	12,1	7,8	-4,3	-35,4
1996 – 1999	15,0	9,9	-5,2	-34,4

1) Betrachtet wurden Personen, die im Zeitraum 1996 bis 1998, 1996 bis 1999 bzw. 1996 bis 2000 in einem Privathaushalt in Deutschland gelebt haben, im Jahr 1996 im Alter zwischen 18 und 65 Jahren waren und im April 1996 einer Erwerbstätigkeit nachgingen. Die Werte in den Spalten 3 und 4 wurden auf Basis der ungerundeten Werte berechnet. – 2) Analyse mit Daten, die für jeden Monat im Betrachtungszeitraum den Erwerbsstatus ausweisen. – 3) Analyse mit Statusvariablen, die den Erwerbsstatus für jedes Jahr im April ausweisen. – – – Datenquelle: SOEP (Episodendaten); eigene Berechnungen; hochgerechnete Ergebnisse

4.2.2 Beschränkte Möglichkeiten der Bestimmung von Verweildauern

Die Konzeption des Mikrozensus bedingt, dass sich das Mikrozensus-Panel nur sehr begrenzt für die Analyse von Verweildauern eignet. Es soll an dieser Stelle jedoch auf zwei Möglichkeiten hingewiesen werden: Im Mikrozensus-Längsschnitt 1996 bis 1999 ist in allen vier Erhebungsjahren für Erwerbstätige die Information verfügbar, wann die Tätigkeit beim aktuellen Arbeitgeber aufgenommen wurde (Jahr und Monat). Arbeitssuchende wurden in allen vier Jahren gefragt, wie lange ihre Arbeitssuche schon andauert. Bei Arbeitssuchenden ist außerdem die Angabe

vorhanden, was vor Beginn der Arbeitssuche gemacht wurde. Diese Informationen eröffnen zum einen die Möglichkeit, die Verweildauer bis zu einem Übergang in Erwerbstätigkeit von denjenigen Personen in den Blick zu nehmen, die im Ausgangsjahr erwerbslos waren.²³⁾ Zum anderen könnte analysiert werden, wovon es abhängt, wie viel Zeit bis zu einem Wechsel des Arbeitgebers vergeht. In beiden Fällen müssen verschiedene Annahmen getroffen werden und es wäre zu prüfen, inwiefern der Sachverhalt, dass die Verweildauer bei vielen Personen nur näherungsweise bestimmt werden kann, Probleme aufwirft.²⁴⁾ Ein Potenzial für Analysen ist jedoch in jedem Fall vorhanden.

5 Zusammenfassung und abschließende Bemerkungen

Der vorliegende Beitrag hatte das Ziel, darüber zu informieren, welche Möglichkeiten das Mikrozensus-Panel für die Analyse von Arbeitsmarktübergängen bietet. Gezeigt wurde, dass prinzipiell ein großes Potenzial vorhanden ist, zwei Problemen jedoch Beachtung geschenkt werden muss.

Das erste Problem besteht darin, dass Personen, die aus einem Auswahlbezirk fortziehen, beim Mikrozensus nicht wiederbefragt werden und folglich in Längsschnittanalysen nicht einbezogen werden können. Bei der Bestimmung von Übergangswahrscheinlichkeiten führt dies in einigen Fällen zu substantziellen Verzerrungen. Das Problem wird dadurch entschärft, dass sich in Tests gezeigt hat, dass der Ausfall räumlich mobiler Personen durch Gewichtung in hohem Maße kompensiert werden kann. Weiterhin scheint es so zu sein, dass multivariate Analysen von der Problematik nur in relativ geringem Maße betroffen sind.

Das zweite Problem ist der Sachverhalt, dass das Mikrozensus-Panel nur vereinzelt Retrospektiv- bzw. Verlaufsinformationen bereitstellt. Informationen, über die Zeiten zwischen den einzelnen Mikrozensus-Berichtswochen sowie Informationen über die Zeit vor der ersten Berichtswoche in einem Längsschnitt, sind aus diesem Grund in der Regel nicht verfügbar. Zum einen ergibt sich hieraus eine Problematik unbeobachteter Übergänge. Das Berichtswochenkonzept führt zum anderen dazu, dass die Möglichkeiten der Betrachtung von Verweildauern sehr begrenzt sind. Auf folgende Punkte kann jedoch verwiesen werden: Erstens lässt sich z. B. mit Daten des SOEP relativ gut quantifizieren, wie groß bestimmte Verzerrungen ausfallen, die sich aus unbeobachteten Übergängen ergeben. Zweitens dürfte auch für unbeobachtete Übergänge gelten, dass multivariate Analysen von der Problematik insgesamt nur in relativ geringem Maße betroffen sind. Drittens sind unbeobachtete Übergänge nur in bestimmten Fällen ein mögliches Problem, weil unter Umständen nur von Interesse ist, welche Status von Personen zu einzelnen Erhebungszeitpunkten eingenommen wurden. Viertens ist schließlich darauf zu verweisen, dass trotz spärlicher Informationen einzelne Möglichkeiten der Analyse von Verweildauern bestehen. Insgesamt lässt sich aus diesem Grund als Fazit festhalten, dass sich aus den zwei diskutierten Problemen einige Beschränkungen ergeben, dass das Analysepotenzial des Mikrozensus-Panels durch die aufgezeigten Probleme jedoch nicht grundsätzlich beeinträchtigt wird.

²³⁾ Als „erwerbslos“ müssten in diesem Fall Personen betrachtet werden, die nicht erwerbstätig sind und gemäß Selbstausskunft (aktiv) nach einer Beschäftigung suchen. – ²⁴⁾ Sofern Unterschiede zwischen Vergleichsgruppen von Interesse sind, ist dies möglicherweise nicht schwerwiegend.

Literatur

Basic, Edin; Marek, Ivo; Rendtel, Ulrich (2005): The German Microcensus as a Tool for Longitudinal Data Analysis: An Evaluation Using SOEP Data. In: Schmollers Jahrbuch 125 (2005), S. 1 – 16.

Frick, Joachim (1996): Lebenslagen im Wandel: Determinanten kleinräumiger Mobilität in Westdeutschland. Frankfurt; New York: Campus.

Heidenreich, Hans-Joachim (2002): Längsschnittdaten aus dem Mikrozensus: Basis für neue Analysemöglichkeiten. In: Allgemeines Statistisches Archiv, Bd. 86, Heft 2, S. 213 – 231.

Jürges, Hendrik (2005): Retrospective Error in SOEP Calendar Data: The Case of Unemployment. In: Schmollers Jahrbuch 125 (2005), S. 157 – 165.

Konold, Michael (2005a): Ausmaß und Folgen unbeobachteter Übergänge im Rahmen des Mikrozensus-Panels: Ergebnisse empirischer Analysen. In: Statistische Analysen und Studien Nordrhein-Westfalen, Band 23, S. 22 – 33.

Konold, Michael (2005b): Der Ausfall räumlich mobiler Personen im Mikrozensus-Panel und seine Auswirkungen auf Übergangsanalysen im Familien- und Erwerbsbereich. Ergebnisse empirischer Untersuchungen. Unveröffentlichtes Arbeitspapier. Düsseldorf.

Möhring, Katja; Zühlke, Sylvia (2005): Räumliche Mobilität und Übergänge von Erwerbslosigkeit in Erwerbstätigkeit. In: Statistische Analysen und Studien Nordrhein-Westfalen, Band 24, S. 13 – 23.

Schmidt, Simone (2000): Erwerbstätigkeit im Mikrozensus. Konzept, Definition, Umsetzung. ZUMA-Arbeitsbericht, Nr. 2000 – 01.

Stauder, Johannes (2003): Räumliche Mobilität und Familienzyklus – eine Analyse auf der Basis des Sozio-oekonomischen Panels. In: Statistische Analysen und Studien Nordrhein-Westfalen, Band 11, S. 3 – 12.

Zühlke, Sylvia (2001): Längsschnittanalysen auf der Basis des Mikrozensus. Methodische Probleme und Lösungsansätze. In: Statistische Analysen und Studien Nordrhein-Westfalen, Band 4, S. 3 – 13.

Zühlke, Sylvia (2003): Systematische Ausfälle im Mikrozensus-Panel: Ausmaß und Auswirkungen auf die Qualität von Arbeitsmarktanalysen. In: Allgemeines Statistisches Archiv, Bd. 87, Heft 1, S. 39 – 58.

Autoren- und Kontaktangaben:

Michael Konold, FDZ der Statistischen Landesämter, Landesamt für Datenverarbeitung und Statistik NRW, E-Mail: michael.konold@lds.nrw.de

Analysen regionalspezifischer Verhaltensmuster der Wohnungsnachfrage

Irene Iwanow

Die Übertragung globaler demografischer Trends auf die Einschätzung der regionalen Bevölkerungsentwicklung ist erfahrungsgemäß problematisch und führt für manche Regionen sogar zu Fehlschlüssen hinsichtlich der zu erwartenden Wohnungsmarktentwicklung. Für eine genauere Einschätzung der Auswirkungen des demografischen Wandels auf die Wohnungsmarkt- und Siedlungsflächenentwicklung sind deshalb stärker detaillierte Analysen und Modelle notwendig. Ein Hindernis für differenzierte Analysen ist nicht nur der stark ansteigende Zeitaufwand für Spezialanalysen, sondern insbesondere die fehlende Verfügbarkeit von Regionaldaten für Detailanalysen in den Regionen. Die Schaffung der Forschungsdatenzentren bietet den Nutzern der statistischen Daten eine interessante Möglichkeit, um Detailanalysen aus den vorhandenen Mikrodaten der statistischen Landesämter selbst zu erstellen. Die zur Verfügung stehenden Merkmale können dabei auf verschiedene Weise kombiniert werden, bis sich zum Beispiel die regionalspezifischen Verhaltensmuster der Wohnpräferenzen und der Wohnungsnachfrage abbilden lassen. Am Beispiel verschiedener Analysen zur Entwicklung der Wohneigentumsquote wird gezeigt, wie sich diese für einzelne Haushaltstypen sowie für unterschiedliche Gemeindegrößenklassen unterscheidet. Eigene Analysen im Mikrozensus verschiedener Jahre und in den Gebäude- und Wohnungsstichproben bieten die Grundlage dafür, dass für regional differenzierte Prognosen der Wohnungsnachfrageentwicklung die Wohnsituation der Haushalte im Basisjahr einer Prognose regional differenziert geschätzt werden kann, auch wenn keine regionalspezifischen Auswertungen auf der Gemeindeebene möglich sind.

1 Einführung

Seit einiger Zeit richtet sich das Interesse der wissenschaftlichen Fachöffentlichkeit verstärkt auf Analysen zum demografischen Wandel. Neben Fragen der Auswirkung von Alterung und Schrumpfung auf das Rentensystem und den Arbeitsmarkt erscheinen insbesondere Fragen zur zukünftigen Wohnungsmarkt- und Regionalentwicklung zunehmend Beachtung zu finden. Auch die Statistische Woche 2005, eine gemeinsame Fachtagung der Deutschen Städtestatistiker, der Deutschen Statistischen Gesellschaft und der Deutschen Gesellschaft für Demographie fand unter dem Generalthema „Was passiert, wenn nichts passiert? Folgen des demografischen Wandels für die Stadtgesellschaft“ statt. In den ostdeutschen Kommunen wurden die Auswirkungen der demografischen Veränderungen bereits in den 1990er Jahren immer deutlicher. Der Trend der alternden Bevölkerung wird in Ostdeutschland zusätzlich durch die kontinuierlichen Abwanderungen verstärkt. Die Mitte bis Ende der 1990er Jahre schnell wachsende Zahl von Wohnungsleerständen in allen ostdeutschen Kommunen machte deutlich, dass sich am ostdeutschen Wohnungsmarkt eine in Deutschland noch nicht da gewesene Entwicklung herausbildet. Bislang wurden in der früheren Bundesrepublik Phasen von Wohnungsüberangeboten stets durch Phasen der Wohnungsknappheit abgelöst. Deshalb mussten sich die Kommunalvertreter wenig Sorgen darüber machen, wer die Wohnungsbestände zukünftig nutzen wird. Obwohl das

Thema Alterung ein regionsübergreifender demografischer Faktor ist, welcher in die Kommunal- und Regionalplanung einzubeziehen ist, können auf Grund unterschiedlicher Migrationsmuster in den einzelnen Regionen keine allgemein gültigen Trends für den Wohnungs- und Immobilienmarkt sowie die zukünftige Flächenentwicklung abgeleitet werden. Sowohl die Wohnungspolitik als auch die Kommunal- und Regionalplanung benötigen jedoch zuverlässige Informationen, welche Auswirkungen die demografische Entwicklung in ihrem Einflussbereich haben wird. Mangels vorhandener kleinräumiger Vorausberechnungen können anstehende Entscheidungen meist nicht auf der Basis von Regionalprognosen getroffen werden, sondern müssen aus Landes- oder Bundesprognosen abgeleitet werden. Die Ursachen für fehlende kleinräumige Prognosen liegen in der nicht vorhandenen Datengrundlage zur Haushaltsentwicklung und zur Wohnsituation der Haushalte auf der Kreis- und Gemeindeebene. Auf der kommunalen Ebene ist die Datengrundlage kaum besser und zudem meist uneinheitlich, so dass Prognosen, die über die kommunale Ebene hinaus reichen, ebenfalls nicht leistbar sind.

2 Forschungsziel

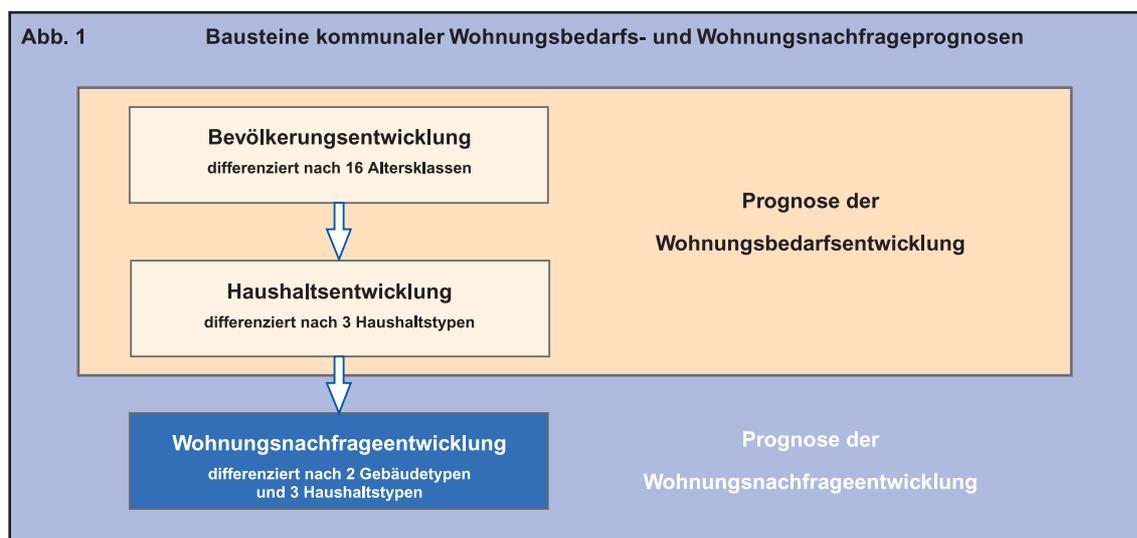
Obwohl der Bedarf an Vorausberechnungen der Wohnungsmarktentwicklungen hoch ist, werden gerade auf der kleinräumigen Ebene noch relativ selten Wohnungsnachfrageprognosen und Prognosen zur Einschätzung der zukünftigen Siedlungsflächenentwicklung sowie des Wohnbaulandbedarfs erstellt. Einerseits fehlen, wie bereits erwähnt, die entsprechenden regionalisierten Ausgangsinformationen zur Haushalts- und Wohnungsmarktentwicklung, andererseits stehen auf dem Gebiet kleinräumiger Prognosen auch kaum methodische Konzepte zur Verfügung, welche die Auswirkungen von Bevölkerungsschrumpfung erfassen können. So müssen einerseits die methodischen Ansätze überarbeitet werden und andererseits muss versucht werden, spezifische Regionalinformationen zu gewinnen. Mit dem bisherigen statistischen Erfassungssystem ist dies aber nicht möglich, denn für gute Regionalinformationen wären Vollerfassungen auf dem Gebiet der Bevölkerungs- und Haushaltsentwicklung sowie zur Wohnsituation der Haushalte notwendig. Da diese auch in den nächsten Jahren nicht zur Verfügung stehen werden, müssen die entsprechenden notwendigen Regionalinformationen aus den vorhandenen statistischen Erhebungen geschätzt werden. Für die methodische Weiterarbeit an Regionalprognosen ist es wichtig, die methodische Arbeit an den Modellen und die statistische Schätzung regionaler Grundlagen gut aufeinander abzustimmen.

Durch die Einrichtung der Forschungsdatenzentren tun sich gerade für den analytischen Bereich in der Regionalforschung ganz neue Möglichkeiten auf. Während es im Bereich der Bevölkerungsstatistik schon immer eine gute regionalisierte Datenbasis gab, beruhen die regionalen Haushaltsangaben auf dem jährlich durchgeführten Mikrozensus, welcher als 1 %ige Stichprobe durchgeführt wird und aus diesem Grunde auch nur begrenzt regionalisierbar ist. Weitere Informationen zur Wohnsituation der Haushalte werden nur in den Zusatzerhebungen zum Mikrozensus oder in den 1 %igen Gebäude- und Wohnungsstichproben der vergangenen Jahre erhoben. Eigene Analysen zur Wohnsituation ausgewählter Haushaltsgruppen, die in den regulären statistischen Berichten auf der Basis des Mikrozensus nicht üblich sind, lassen sich über die Forschungsdatenzentren kostengünstig realisieren. Um bei diesen Analysen möglichst geringe Verluste bei den Fallzahlen zu erreichen, wurde die Variante des kontrollierten Fernrechnens genutzt.

In dem vom Leibniz-Institut für ökologische Raumentwicklung e. V. (IÖR) bearbeiteten Projekt „Regionalisierte Wohnungsnachfrageprognose für Baden-Württemberg“ im Auftrag der Arbeitsgemeinschaft Baden-Württembergischer Bausparkassen werden auf der Basis des von Iwanow (2002) entwickelten methodischen Ansatzes für kommunale Wohnungsnachfrageprognosen kleinräumige Wohnungsmarktprognosen auf der kommunalen Ebene erstellt. Das Ziel dieser Studie bestand darin, die regionalen Disparitäten der demografischen Entwicklung in Baden-Württemberg zu erkennen und ihre Auswirkungen auf die Entwicklung der Wohnungsnachfrage abzuschätzen sowie das daraus resultierende regionale Nachfragepotenzial nach Wohnungsneubau zu quantifizieren. Ähnliche Analysen und Prognosen werden auch für Gemeinden aus Sachsen im Projekt „Grundlagen regional differenzierter Siedlungsflächenprognosen“ durchgeführt. In diesem Projekt schließt sich noch eine Abschätzung des zu erwartenden Nachfragepotenzials nach Wohnbauland an (Horeni, 2005).

3 Methodischer Aufbau der Nachfrageprognosen

Die Prognose der kommunalen Nachfrageentwicklung basiert auf Vorausschätzungen der Bevölkerungs- und Haushaltentwicklung (Abb. 1). Auf der Grundlage einer Haushaltsprognose kann zunächst nur der zu erwartende Bedarf an Wohnungen quantifiziert werden, der notwendig ist, damit jeder Haushalt auch mit einer Wohnung versorgt werden kann. So bleiben in Wohnungsbedarfsprognosen unterschiedliche Wohnpräferenzen, wie zum Beispiel die Unterscheidung der Wohnwünsche der Haushalte nach Wohneigentum oder Mietwohnung, noch unberücksichtigt. Eine derartige Differenzierung der Wohnpräferenzen der Haushalte nach verschiedenen Wohnformen wird erst im Baustein „Nachfrageprognose“ vorgenommen. Das methodische Konzept der kommunalen Nachfrageprognose des Leibniz-Instituts für ökologische Raumentwicklung e. V. sieht vor, dass Differenzierungen nach maximal sieben Gebäudetypen bzw. Wohnungsteilmärkten möglich ist. Bei der Bearbeitung von Regionalprognosen, die sich auf mehrere Kommunen beziehen, erscheint eine Differenzierung nach zwei Gebäudetypen, die Ein- und Zweifamilienhäuser und die Mehrfamilienhäuser, maximal leistbar. Weitere Differenzierungen der Gebäudestrukturen unterscheiden sich in der Regel von Kommune zu Kommune deutlich, sodass eine einheitliche Strukturierung nicht möglich ist.



Quelle: Darstellung in Anlehnung an Iwanow (2002)

Für die Bearbeitung kommunaler Wohnungsnachfrageprognosen sind somit drei aufeinander aufbauende Bausteine notwendig. Die beiden Bausteine „Bevölkerungsentwicklung“ und „Haushaltsentwicklung“ werden dabei nur kurz erläutert. Die durchgeführten Regionalanalysen und -schätzungen beziehen sich jedoch vor allem auf den Baustein „Nachfrageentwicklung“, der bei den weitergehenden Betrachtungen im Zentrum steht.

Im Baustein „Bevölkerungsentwicklung“ ist auf kommunaler Ebene eine Vorausschätzung der zu erwartenden Einwohnerzahl, welche zukünftig mit Wohnraum zu versorgen ist, möglich. Damit dieser Baustein auch für Kleinstädte und größere Gemeinden nutzbar ist, wird im demografischen Teil der Nachfrageprognose die Bevölkerung nur nach 16 Altersklassen differenziert.

Ergänzt durch Annahmen zu den zukünftigen Zu- und Abwanderungen und unter Beachtung ihrer Altersstruktur werden die räumlichen Bevölkerungsbewegungen prognostiziert. Hinzu kommt noch der biometrische Teil der Bevölkerungsprognose, die Prognose der Zahl der Geburten und der zu erwartenden Sterbefälle. Dabei wird in der Nachfrageprognose des IÖR von dynamisierten Geburten- und Sterberaten ausgegangen. Die erforderlichen Datengrundlagen für die Prognose der Bevölkerungsentwicklung liegen sowohl bei der Kommunalstatistik als auch bei den statistischen Landesämtern ausreichend detailliert vor.

Problematisch sieht es hingegen bei der Datengrundlage für eine Vorausberechnung der Haushaltsentwicklung aus. Auf der kommunalen Ebene gibt es nur in einigen Kommunen Informationen über die Zahl der in der Kommune lebenden Haushalte. Einige Städte machen auf der Basis der beim KOSIS-Verbund erhältlichen Software HHGEN¹⁾ eigene Haushaltsschätzungen. Einige statistische Landesämter werten den Mikrozensus auch bis zur Kreisebene aus, sodass in diesen Ländern dann für die kreisfreien Städte Informationen zur kommunalen Haushaltsentwicklung vorliegen. Eine Regionalisierung des Mikrozensus bis zur Kreisebene nehmen jedoch nicht alle statistischen Landesämter vor. In kleinräumigen Regionalprognosen sind deshalb für die einzelnen Städte und Gemeinden Schätzungen zur Zahl und Struktur der Haushalte im Basisjahr der Prognose notwendig. Für die Prognosejahre selbst wird die Haushaltsentwicklung dann aufbauend auf der prognostizierten Bevölkerungsentwicklung sowie der zu erwartenden Veränderungen im Haushaltsbildungsverhalten prognostiziert. Für die sich anschließende Vorausschätzung der Nachfrageentwicklung ist insbesondere eine Differenzierung der Haushalte nach der Haushaltsgröße und dem Alter der Haushaltsmitglieder von Interesse.

Bei einer Differenzierung der Nachfrageentwicklung nach Haushaltstypen sind vor allem jene interessant, die sich hinsichtlich ihrer Wohnpräferenzen sowie ihrer Aktivitäten am Wohnungsmarkt unterscheiden. Ein für die zukünftige Nachfrageentwicklung besonders wichtiger Haushaltstyp sind die „Jüngeren Ein- und Zwei-Personen-Haushalte“. Zu diesem Haushaltstyp gehören insbesondere die Haushaltsgründer sowie die jungen Haushalte, deren Wohnansprüche sich noch relativ schnell verändern. Diese Haushalte sind als Nachfrager am Wohnungsmarkt besonders aktiv und zusätzlich ist ihre Wohndauer in einer Wohnung meist weniger lange als bei älteren oder größeren Haushalten. Ein zweiter Haushaltstyp umfasst die „Haushalte mit drei und mehr Personen“. Zu diesem Haushaltstyp gehören insbesondere die Familienhaushalte, deren Interesse an der Bildung von Wohneigentum i. Allg. besonders hoch ist. Die Wohnpräfe-

1) KOSIS-Verbund: Koordinierte Haushalte- und Bevölkerungsstatistik aus dem Melderegister. <http://www.kosis.de>

renzen dieses Haushaltstyps sind im Vergleich zu den „Jüngeren Ein- und Zwei-Personen-Haushalten“ stärker durch qualitative Wohnwünsche geprägt. Neben dem Erwerb von Wohneigentum gehört insbesondere der Wunsch nach einer größeren Wohnung oder nach einem besseren Wohnumfeld zu den typischen qualitativen Wohnwünschen. Als dritter Haushaltstyp werden die „Älteren Ein- und Zwei-Personen-Haushalte“ betrachtet. Die Wohnpräferenzen dieses Haushaltstyps beziehen sich insbesondere auf Verbesserungen des Wohnkomforts. Verbesserungen der Wohnflächenausstattungen werden oft auf eine passive Weise, wie z. B. durch den Auszug der Kinder, realisiert. Haushalte dieses Haushaltstyps sind insgesamt weniger mobil als die beiden anderen Gruppen, da sie ihre Wohnwünsche im Laufe ihrer Haushaltsbiographie meist schon recht gut verwirklichen konnten.

Die Vorausschätzung und Differenzierung der Haushalte nach den drei Haushaltstypen basiert einerseits auf Parametern aus der altersstrukturellen Bevölkerungsentwicklung und andererseits auf der Annahme zur weiteren Verkleinerung der Haushaltsgröße. Die Dynamik der Haushaltsverkleinerung wird dabei insbesondere über die durchschnittliche Haushaltsgröße der einzelnen Altersgruppen der Bevölkerung sowie ihrer quantitativen altersstrukturellen Veränderungen geschätzt. Eine Schätzung des Anteils der Ein- und Zwei-Personen-Haushalte beruht zusätzlich auf einer einfachen Regressionsgleichung, welche den Zusammenhang zwischen der durchschnittlichen Haushaltsgröße und dem Anteil der Ein- und Zwei-Personen-Haushalte beschreibt. Eine Differenzierung in jüngere und ältere Ein- und Zwei-Personen-Haushalte erfolgt über die altersstrukturelle Differenzierung der Bevölkerung.

Die Basis für den Prognosebaustein „Nachfrageentwicklung“ bildet eine Differenzierung der Haushalte nach unterschiedlichen Wohnpräferenzen. Hinsichtlich der Wohnsituation der Haushalte liegen auf der kommunalen Ebene keine Informationen mehr vor. Die einzigen vorhandenen Datenquellen sind wiederum der Mikrozensus oder die Gebäude- und Wohnungsstichprobe in ihrer regionalen Differenzierung bis zur Ebene der Raumordnungsregionen. Hinzu kommen eventuell spezifische Bewohnerbefragungen kommunaler Wohnungsgesellschaften oder Wohnungsgenossenschaften, welche jedoch für andere Kommunen eine zu geringe Aussagekraft haben. Somit muss, wie bei der Haushaltsvorausschätzung, die Wohnsituation der Haushalte auch für das Basisjahr der Prognose mit Hilfe statistischer Methoden aus der amtlichen Statistik geschätzt werden. So ist zunächst, trotz aller noch vorhandenen methodischen Mängel, eine annähernde Beschreibung der Wohnsituation der Haushalte im Basisjahr der Prognose möglich. Diese grobe Schätzung der Wohnsituation der Haushalte wird dann als Nachfrage der Haushalte im Basisjahr interpretiert, denn jeder Haushalt fragt seine Wohnung, in der er derzeit wohnt, solange weiter nach, bis er in eine andere Wohnung oder ins Eigenheim umziehen möchte. In diesem Sinne wird auch jeder Haushalt als nachfragender Haushalt interpretiert. Eine Darstellung der kleinräumigen Nachfrageanalysen wird im Abschnitt 4 gegeben.

Die Vorausschätzung der Nachfrageentwicklung für die einzelnen Prognosejahre setzt auf der Beschreibung der Nachfragesituation im Basisjahr der Prognose auf. Aufbauend auf dieser werden für den Prognosezeitraum die Entscheidungen der Haushalte zur Veränderung ihrer Wohnsituation bzw. für das weitere Bewohnen ihrer jetzigen Wohnung simuliert. Als Basis für diese Entscheidungen der Haushalte dienen die auf einer Mikrolängsschnittbeobachtung von Haushalten analysierten Umzugs- und Bleibewahrscheinlichkeiten der einzelnen Haushaltstypen. Während die Umzugswahrscheinlichkeiten entweder den Wechsel von einer Mietwohnung ins

Wohneigentum oder umgekehrt beschreiben, geben die Bleibewahrscheinlichkeiten an, wie hoch der Anteil der Haushalte ist, die entweder in ihrer Wohnung wohnen bleiben oder innerhalb des gleichen Marktsegmentes (Mietwohnung oder Wohneigentum) umziehen. Mit Hilfe der Umzugswahrscheinlichkeiten kann ein stochastischer Entscheidungsprozess der Haushalte hinsichtlich ihrer Wohnungswahl nachgebildet werden. Als Datenbasis für die Berechnung der Umzugs- und Bleibewahrscheinlichkeiten dienen Analysen auf der Basis des Sozio-ökonomischen Panels (SOEP),²⁾ einer vom Deutschen Institut für Wirtschaftsforschung betreuten, jährlich wiederkehrenden, repräsentativen Personen- und Haushaltsbefragung. Eine grobe Differenzierung der Umzugswahrscheinlichkeiten nach der Gemeindegroße ist dabei möglich. Diese Analysen auf der Basis des SOEP³⁾ können nicht über das Forschungsdatenzentrum durchgeführt werden und sind deshalb nicht Gegenstand dieses Beitrags.

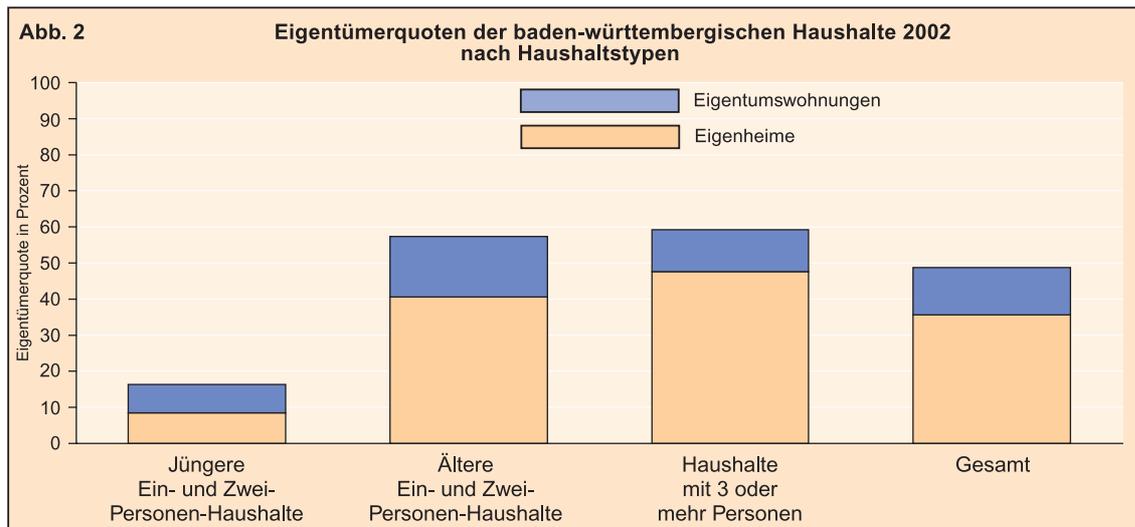
4 Analysen zur Wohnsituation der Haushalte

Als statistische Basis für die Entwicklung und Schätzung der Wohnsituation der Haushalte werden die Gebäude- und Wohnungsstichprobe 1993 sowie die Zusatzerhebungen zum Mikrozensus 1998 und 2002 herangezogen. Die einzelnen Statistiken können dabei nicht nur nach den in den Berichten der statistischen Landesämter oder des Statistischen Bundesamtes üblichen Differenzierungen vorgenommen werden. Für eigene Analysen auf der Basis des Mikrozensus und der Gebäude- und Wohnungsstichprobe über die Forschungsdatenzentren sind auch eigene Klassifizierungsmerkmale möglich. Bei detaillierten Analysen gilt es jedoch insbesondere darauf zu achten, dass die Fallzahlen hinsichtlich der einzelnen Klassifizierungsmerkmale und vor allem bei der Kombination mehrerer Merkmale nicht zu gering werden, sodass eine entsprechende Repräsentativität der Stichprobenergebnisse noch gegeben ist.

Für die Schätzung der Wohnsituation der baden-württembergischen Haushalte in den einzelnen Kommunen im Jahr 2001, dem Basisjahr der zu erarbeitenden Nachfrageprognose für Baden-Württemberg, war es notwendig, die Wohnsituation der drei Haushaltstypen im Mikrozensus 2002 so genau wie möglich zu analysieren. Da die Mikrozensus-Befragung bislang im April des jeweiligen Jahres erfolgte, war eine relativ zeitnahe Beschreibung der Wohnsituation möglich. Obwohl die Wohnsituation der baden-württembergischen Haushalte auch auf der Basis der Gebäude- und Wohnungsstichprobe 1993 sowie des Mikrozensus 1998 analysiert wurde, werden aus Gründen der Übersichtlichkeit vorwiegend die Ergebnisse aus dem Mikrozensus 2002 dargestellt. Vergleiche mit früheren Befragungen zur Wohnsituation der Haushalte ergaben analoge Wohnmuster. Stellvertretend für die möglichen Differenzierungen der Wohnsituation der Haushalte werden in diesem Beitrag die Analysen zur Eigentümerquote vorgestellt. Die Eigentümerquote beschreibt dabei den Anteil der im selbst genutzten Wohneigentum wohnenden Haushalte zur Gesamtzahl der Haushalte. Für die Nachfrageprognose werden die im Wohneigentum lebenden Haushalte zusätzlich noch danach differenziert, ob sie im Eigenheim oder in einer Eigentumswohnung wohnen.

2) Jährlich durchgeführte Personen- und Haushaltsbefragung seit 1984 zur Gewinnung von Mikrodaten für die sozial- und wirtschaftswissenschaftliche Grundlagenforschung. – 3) Analysen auf der Basis des SOEP für die neuen Länder siehe Iwanow (1998), Bundesamt für Bauwesen und Raumordnung (2001).

Die Analysen zum Mikrozensus 2002 spiegeln beispielsweise für Baden-Württemberg sehr charakteristische Wohnmuster für alle drei Haushaltstypen wider (Abb. 2). Der höchste Anteil der im Wohneigentum wohnenden Haushalte ist erwartungsgemäß bei den „Haushalten mit drei oder mehr Personen“ zu erkennen. So lag die Eigentümerquote in diesem Haushaltstyp in Baden-Württemberg im Jahr 2002 bei 59,2 %. Einen nur gering darunter liegenden Wert weist die Gruppe der „Älteren Ein- und Zwei-Personen-Haushalte“ auf. Von den Haushalten, die im Jahr 2002 zu dieser Gruppe gehörten, wohnten 57,3 % im selbst genutzten Wohneigentum. Dabei lag der Anteil der Haushalte, die in einer Eigentumswohnung lebten, um mehr als fünf Prozentpunkte höher als in der Gruppe der „Haushalte mit drei oder mehr Personen“.

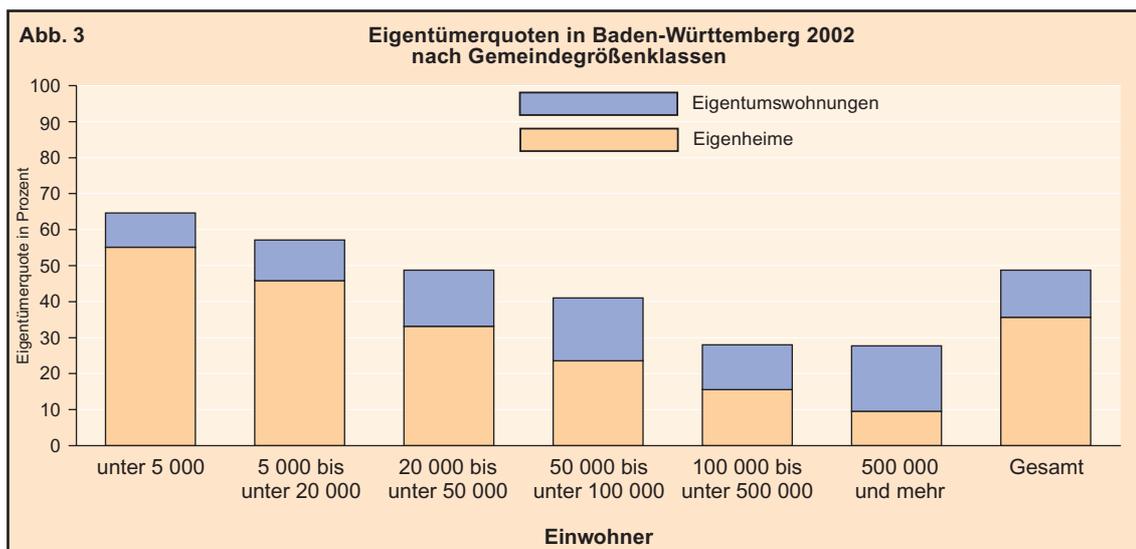


Quelle: Analysen des IÖR, Datenbasis: Zusatzerhebung zum Mikrozensus 2002

Die analysierten Verhaltensmuster beider Haushaltsgruppen entsprechen den Erwartungen, denn erfahrungsgemäß wohnen kleinere Haushalte häufiger als große Haushalte in Eigentumswohnungen und umgekehrt leben die klassischen Familienhaushalte, repräsentiert durch die „Haushalte mit drei oder mehr Personen“ häufiger im selbst genutzten Eigenheim als dies bei den „Älteren Ein- und Zwei-Personen-Haushalten“ der Fall ist. Mit zunehmendem Alter wechseln Eigenheimbesitzer manchmal auch in eine Eigentums- oder sogar in eine Mietwohnung mit hohem Wohnkomfort. Die Eigentümerquote der „Jüngeren Ein- und Zwei-Personen-Haushalte“ ist mit 16,3 % relativ niedrig. Daraus ist ersichtlich, dass die meisten Haushalte erst in der Familienphase oder in einer späteren Lebensphase ihre Wohnpräferenz für das Wohneigentum entwickeln und in dieses umziehen. Hinzu kommen bei den „Jüngeren Ein- und Zwei-Personen-Haushalten“ insbesondere auch finanzielle Restriktionen, welche die Umsetzung des Wunsches nach Wohneigentum noch nicht zulassen.

Neben den unterschiedlichen Wohnpräferenzen der einzelnen Haushaltstypen gibt es im Wohnverhalten auch typische siedlungsstrukturelle Unterschiede, die sich analytisch insbesondere durch Differenzierungen nach der Gemeindegröße abbilden lassen. So wohnen zum Beispiel traditionell in den ländlichen Gemeinden mehr Haushalte im selbst genutzten Wohneigentum als dies in den städtisch geprägten Kommunen der Fall ist (Abb. 3).

In den Gemeinden mit weniger als 5 000 Einwohnern wohnten 2002 durchschnittlich 64,6 % der Haushalte im selbst genutzten Wohneigentum, davon rund 85 % in einem Eigenheim, die restlichen Wohnungseigentümer in Eigentumswohnungen. Mit zunehmender Gemeindegröße und



damit auch steigendem Verstädterungsgrad nimmt i. Allg. der Anteil der im Wohneigentum lebenden Haushalte ab. So wohnten 2002 in den Gemeinden mit mindestens 100 000 und weniger als 500 000 Einwohnern nur noch 28,0 % der Haushalte im Wohneigentum. Die restlichen 72,0 % der baden-württembergischen Haushalte in Großstädten unter 500 000 Einwohnern wohnten zur Miete. In den Großstädten mit mindestens 500 000 Einwohnern ist der Anteil der Eigentümerhaushalte nur wenig geringer als in den Großstädten mit weniger als 500 000 Einwohnern. Unterschiede gibt es bei den beiden Gruppen der Großstädte aber hinsichtlich der Differenzierung nach Eigentumswohnungen und Eigenheimen. In den Großstädten mit mindestens 500 000 Einwohnern, in Baden-Württemberg trifft dies nur auf Stuttgart zu, wohnen deutlich mehr Haushalte in Eigentumswohnungen (18,2 %) als in Eigenheimen (9,5 %). Im Gegensatz dazu wohnen in den Großstädten mit weniger als 500 000 Einwohnern nur 12,5 % der Haushalte in Eigentumswohnungen, aber 15,5 % in Eigenheimen. Allgemein lässt sich erkennen, dass mit zunehmendem Verstädterungsgrad der Anteil der in Eigenheimen wohnenden Haushalte kontinuierlich abnimmt und der Anteil der in Eigentumswohnungen wohnenden Haushalte ansteigt.

Wie relativ problemlos sichtbar ist, kann bei der Erstellung von Nachfrageprognosen weder auf die Differenzierung der Nachfrage nach unterschiedlichen Haushaltstypen noch auf die siedlungsstrukturellen Besonderheiten der Nachfrage verzichtet werden. Neben der Abbildung der angegebenen charakteristischen Wohnmuster der Haushalte in den einzelnen Gemeindegrößenklassen (Abb. 3) und den einzelnen Haushaltstypen (Abb. 2) muss auch das unterschiedliche Wohnverhalten der einzelnen Haushaltstypen in den einzelnen Gemeindegrößen in die Nachfrageprognose einbezogen werden. Analytisch gesehen ist dies durch die Kombination beider Determinanten möglich. Als Beispiele für die Regionalisierung der unterschiedlichen Wohnpräferenzen der einzelnen Haushalts- bzw. Nachfragergruppen dienen die Gemeinden mit weniger als 5 000 Einwohnern (Tab. 1) und die Großstädte mit mindestens 100 000 und weniger als 500 000 Einwohnern (Tab. 2).

In den Gemeinden mit weniger als 5 000 Einwohnern wohnen 73,5 % der „Haushalte mit drei oder mehr Personen“ im Wohneigentum (Tab. 1), in den Städten mit mindestens 100 000 und weniger als 500 000 Einwohnern sind es hingegen nur 38,8 % (Tab. 2). In Baden-Württemberg insgesamt liegt die Eigentümerquote der „Haushalte mit drei oder mehr Personen“ bei durchschnittlich 59,2 % (Abb. 2).

1. Anteil der Eigentümer-Haushalte 2002 in Gemeinden mit mindestens 100 000 und weniger als 500 000 Einwohnern				
Anteil der Eigentümer-Haushalte in	Jüngere Ein- und Zwei-Personen-Haushalte	Ältere Ein- und Zwei-Personen-Haushalte	Haushalte mit drei oder mehr Personen	Haushalte insgesamt
Eigenheimen	20,7%	59,1%	65,3%	55,1%
Eigentumswohnungen	8,2%	11,2%	8,2%	9,5%
Insgesamt	28,9%	70,3%	73,5%	64,6%

Quelle: Analysen des IÖR, Datenbasis: Zusatzerhebung zum Mikrozensus 2002

Weitere typische Verhaltensmuster bezüglich der Wohnpräferenzen der Haushalte lassen sich auch bei der Differenzierung nach den beiden Wohnformen Eigenheim oder Eigentumswohnung erkennen. Während in den Großstädten mit weniger als 500 000 Einwohnern beispielsweise der Anteil der in Eigentumswohnungen lebenden „Jüngeren Ein- und Zwei-Personen-Haushalte“ mit 5,2 % mehr als doppelt so hoch wie in den Eigenheimen mit 2,4 % liegt (Tab. 2), zeigt sich in den Gemeinden mit weniger als 5 000 Einwohnern ein völlig anderes Bild. In diesen Gemeinden beträgt der Anteil der in Eigentumswohnungen lebenden „Jüngeren Ein- und Zwei-Personen-Haushalte“ 8,2 % und liegt damit noch nicht einmal bei der Hälfte der „Jüngeren Ein- und Zwei-Personen-Haushalte“, die in Eigenheimen leben. Ihr Anteil betrug in den Gemeinden Baden-Württembergs mit weniger als 5 000 Einwohnern im Jahr 2002 etwa 20,7 % (Tab. 1) und lag wesentlich höher als im Durchschnitt der Gemeinden Baden-Württembergs (Abb. 2). Auf weitere Darstellungen charakteristischer Vergleiche zum Wohnverhalten wird verzichtet. Diese sind aber ähnlich plausibel wie die angeführten Beispiele.

2. Anteil der Eigentümer-Haushalte 2002 in Gemeinden unter 5 000 Einwohnern				
Anteil der Eigentümer-Haushalte in	Jüngere Ein- und Zwei-Personen-Haushalte	Ältere Ein- und Zwei-Personen-Haushalte	Haushalte mit drei oder mehr Personen	Haushalte insgesamt
Eigenheimen	2,4%	19,8%	25,8%	15,5%
Eigentumswohnungen	5,2%	17,6%	13,0%	12,5%
Insgesamt	7,6%	37,4%	38,8%	28,0%

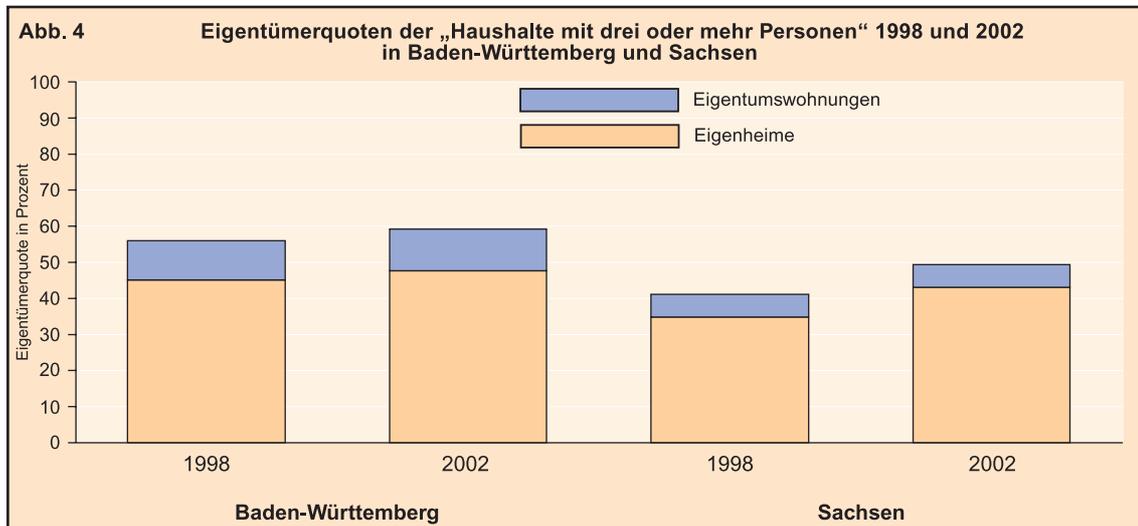
Quelle: Analysen des IÖR, Datenbasis: Zusatzerhebung zum Mikrozensus 2002

Mit Hilfe dieser und weiterer kleinräumiger Differenzierungen des Wohnverhaltens auf der Basis von Sonderauswertungen aus dem Mikrozensus über das Angebot des Forschungsdatenzentrums sind regionalisierte Schätzungen des Wohnverhaltens der baden-württembergischen Haushalte möglich. Auch wenn nicht alle in der Realität wirkenden Differenzierungsmerkmale der Nachfrage abgebildet werden können, erscheinen Schätzungen zur kleinräumigen Nachfrage leistbar und für die einzelnen Gemeinden genauer zu sein, als wenn die Länderergebnisse oder die Angaben zu den Raumordnungsregionen verwendet werden. Die kleinräumige Schätzung der Wohnsituation der baden-württembergischen Haushalte setzt sich dabei aus zwei Komponenten zusammen. Eine Komponente bildet die quantitative Schätzung der Haushalte in jedem Haushaltstyp, die zweite Komponente bezieht sich auf die Schätzung der Wohnpräferenzen der einzelnen Haushaltstypen in der entsprechenden Gemeindegrößenklasse. So wird eine regionalisierte Datenbasis geschaffen, mit deren Hilfe kleinräumige Informationen zur Wohnsituation der Haushalte in den einzelnen Gemeinden im Basisjahr der Prognose bereit gestellt werden können. Eine anschließende Überprüfung und eventuelle Anpassung der Schätzungen an die vom Statistischen Landesamt Baden-Württemberg zur Verfügung gestellten Ergebnisse des Mikrozensus zur Wohnsituation in den einzelnen Regionen ermöglicht einen Ausgleich regions-

spezifischer Besonderheiten, welche durch regionale Sondereinflüsse immer wieder gegeben sein werden. Für die Prognose liegt dann für das Basisjahr eine kleinräumige Beschreibung der Wohnsituation in Baden-Württemberg vor, welche alle derzeit zur Verfügung stehenden statistischen Möglichkeiten ausschöpft.

5 Zeitliche und länderbezogene Vergleiche der Wohnsituation

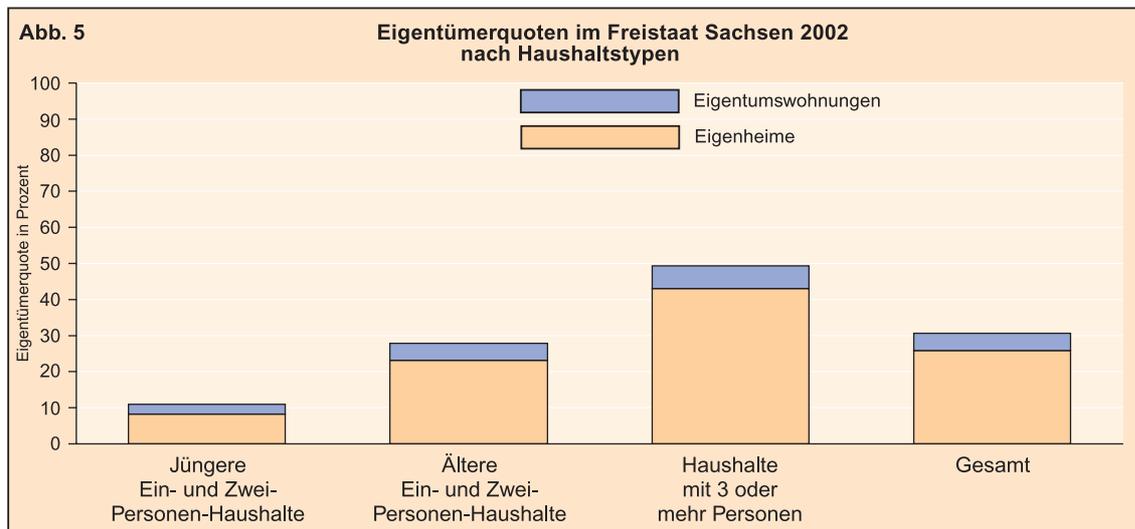
Zeitliche Vergleiche zur Entwicklung der Wohnsituation der Haushalte in der Vergangenheit können zum Beispiel aus der Datenbasis des Mikrozensus 1998 und 2002 abgeleitet werden. So zeigt sich deutlich, dass die Eigentumbildung der Haushalte im Zeitraum 1998 bis 2002 deutlich vorangekommen ist. Am Beispiel der „Haushalte mit drei oder mehr Personen“ ist zu erkennen, dass 2002 mehr Haushalte dieses Haushaltstyps im Wohneigentum wohnen, als dies noch 1998 der Fall war (Abb. 4). In Baden-Württemberg ist bei den „Haushalten mit drei oder mehr Personen“ die Eigentümerquote von 55,9 % im Jahr 1998 auf 59,2 % angestiegen. Das sind 3,6 Prozentpunkte in vier Jahren. 1998 lag die Eigentümerquote der „Haushalte mit drei oder mehr Personen“ im Freistaat Sachsen mit 41,1 % deutlich niedriger als in Baden-Württemberg. In Sachsen hat die Eigentümerquote innerhalb von vier Jahren jedoch erheblich schneller zugenommen. Der Zuwachs betrug im gleichen Zeitraum 8,2 Prozentpunkte. Zusätzlich ist zu erkennen, dass im Gegensatz zu Baden-Württemberg in Sachsen die Zahl der „Haushalte mit drei oder mehr Personen“, die in einer Eigentumswohnung wohnen, nicht zugenommen hat. Im Freistaat Sachsen bezog sich der Eigentumserwerb der Haushalte dieser Haushaltsgruppe fast ausschließlich auf Eigenheime.



Quelle: Analysen des IÖR, Datenbasis: Zusatzerhebung zum Mikrozensus 1998 und 2002

Ein Vergleich der Wohnpräferenzen der drei Haushaltstypen im Freistaat Sachsen mit jenen in Baden-Württemberg lässt teilweise ähnliche, teilweise aber auch deutlich unterschiedliche Verhaltensmuster der Wohnungsnachfrage erkennen. So reichte in Sachsen der Zeitraum nach der politischen Wende noch nicht aus, dass die Gruppe der „Älteren Ein- und Zwei-Personen-Haushalte“ eine annähernd hohe Eigentümerquote wie in Baden-Württemberg erreichen konnte (Abb. 2 und Abb. 5).

Meist haben die heute „Älteren Ein- und Zwei-Personen-Haushalte“ in der Lebensphase, als die Kinder noch im Haushalt lebten, traditionell Wohneigentum gebildet. So ist zu erwarten, dass zukünftig insbesondere bei diesem Haushaltstyp die Eigentümerquote weiter ansteigen wird, auch wenn dieses Wachstum zum großen Teil durch einen passiven Prozess ausgelöst wird. Bei den beiden anderen Haushaltstypen liegen die Eigentümerquoten in Sachsen ebenfalls niedriger als in Baden-Württemberg, jedoch sind die Unterschiede zu Baden-Württemberg geringer als beim Haushaltstyp „Ältere Ein- und Zwei-Personen-Haushalte“.



Quelle: Analysen des IÖR, Datenbasis: Zusatzerhebung zum Mikrozensus 2002

6 Konsequenzen für die Gewinnung von Regionaldaten

Kleinräumige Prognosen haben bislang ihre Begrenzungen in der regionalen Verfügbarkeit von Basisdaten. Diese Hürde war in der Vergangenheit nur schwer zu umgehen, da Sonderauswertungen bei den statistischen Landesämtern oder beim Statistischen Bundesamt sehr kostenintensiv waren. Durch die Möglichkeit, über das Forschungsdatenzentrum eigene Datenanalysen durchführen zu können, sind die Probleme zur regionalen Differenzierung von 1 %-Stichproben zwar insgesamt nicht behoben, aber durch die kostengünstigen Analysen können zumindest eine Reihe von Sonderauswertungen durchgeführt werden, mit denen die Überprüfung der Abbildungsmöglichkeit regionaler Prozesse möglich wird. Die in den Abschnitten 4 und 5 vorgestellten Analysen auf der Basis der Primärdaten der statistischen Landesämter und des Statistischen Bundesamtes zeigen, dass plausible Ergebnisse durch eigene Analysen erzielt werden können. Dabei gilt es jedoch stets darauf zu achten, dass bei der Auswertung der Datenbasis die Fallzahlen nicht zu gering werden. Bei Beachtung dieser Randbedingung können für unterschiedliche Forschungsaufgaben wichtige zusätzliche Informationen gewonnen werden.

Neben den in diesem Beitrag vorgestellten Ergebnissen sind auf der gleichen Datenbasis weitere Analysen zu den Verhaltensmustern der Nachfrager denkbar. So könnten zum Beispiel auch siedlungsstrukturelle Determinanten einbezogen werden. Gleiches gilt auch für haushaltsökonomische Faktoren. Für die statistisch-analytische Praxis würde dies jedoch bedeuten, dass im Sinne der statistischen Sicherheit der Analyseergebnisse andere Differenzierungsmerkmale aufgegeben werden müssten. Trotz der verbesserten Analysemöglichkeiten für wissenschaftliche Arbeiten bleibt festzuhalten, dass auch die besten Analysen zu einzelnen Merkmalen oder einer

Kombination von Merkmalen eine verbesserte regionale Datenbasis oder eine erweiterte Stichprobe bei der Erfassung nicht ersetzen können. Solange sehr wichtige Informationsgrundlagen für die Regionalstatistik nicht vorhanden sind, können jedoch die vorgestellten Analysen als Demonstrationsbeispiel für eine mögliche Behelfslösung dienen. Da sich die regionalen Trends, wie zum Beispiel jene zur Bevölkerungs- und Wohnungsnachfrageentwicklung räumlich immer weiter ausdifferenzieren werden, sollte trotz aller zusätzlichen Analysemöglichkeiten über die Forschungsdatenzentren nicht übersehen werden, dass eine Verbesserung der Grundlagen der Regionalstatistik nicht vernachlässigt werden darf, damit regionale Fehlentscheidungen vermieden werden können.

Literatur

Bundesamt für Bauwesen und Raumordnung (2001): Wohnungsprognose 2015. Bonn = Berichte, 10.

Forschungsdatenzentrum der Länder: Datenfiles des Mikrozensus 1998 und 2002. Statistisches Landesamt Berlin.

Forschungsdatenzentrum der Länder: Datenfile der Gebäude- und Wohnungsstichprobe 1993. Statistisches Landesamt Berlin.

Horeni, S. (2005): Wohnbaulandnachfrageentwicklung unter veränderten Rahmenbedingungen – Prognosen für die Nachfrage nach Ein- und Zweifamilienhäusern am Beispiel der Region Dresden und Großhain, Technische Universität Dresden, Institut für Geographie.

Iwanow, I. (2002): Kommunale Wohnungsnachfrageprognose – eine Methode und Ergebnisse für die Stadt Bautzen. In: Odermatt, A.; Van Wezemael, J. E. (Hrsg.): Geographische Wohnungsmarktforschung – Die Wohnungsmärkte Deutschlands, Österreichs und der Schweiz im Überblick und aktuelle Forschungsberichte, Universität Zürich-Irchel, Geographisches Institut, 97 – 112.

Iwanow, I. (1998): Entwicklung der Wohnflächennachfrage in den neuen Ländern. In: Eichler, K. (Hrsg.): Wohnungsmarktentwicklungen im regionalen Vergleich, Dresden = IÖR-Schriften, 26, 67 – 93.

KOSIS-Verbund: Koordinierte Haushalte- und Bevölkerungsstatistik aus dem Melderegister. <http://www.kosis.de>

Sozio-ökonomisches Panel (SOEP): Mikrodaten der seit 1984 durchgeführten Personen- und Haushaltsbefragung. Deutsches Institut für Wirtschaftsforschung, Berlin.

Autoren- und Kontaktangaben:

Irene Iwanow, Leibniz-Institut für ökologische Raumentwicklung e. V. (IÖR), Weberplatz 1, 01217 Dresden, E-Mail: i.iwanow@ioer.de

Möglichkeiten kleinräumiger Analysen auf Basis des Mikrozensus

Helga Christians

Der vorliegende Beitrag informiert über das Analysepotenzial des Mikrozensus im Bereich kleinräumiger Analysen. Es werden sowohl die wichtigsten methodischen Aspekte des Mikrozensus als auch die im Mikrozensus enthaltenen Gebietseinheiten dargestellt. Auf Grund seines hohen Stichprobenumfangs verfügt der Mikrozensus auf kleinräumiger Ebene noch über verhältnismäßig hohe Fallzahlen und eignet sich für vielfältige inhaltliche Fragestellungen insbesondere im Bereich des Erwerbslebens. Exemplarisch werden im Beitrag die Erwerbsquoten unterschiedlicher Regionen in Nordrhein-Westfalen betrachtet. Darüber hinaus wird dargestellt, welche Nutzungsmöglichkeiten für die Wissenschaft derzeit bereits bestehen und welche für die Zukunft geplant sind.

1 Einleitung

In den Sozial- und Wirtschaftswissenschaften hat es in den letzten Jahren einen erheblichen Bedeutungszuwachs der regionalen Perspektive in wissenschaftlichen Forschungsprojekten gegeben. Im Mittelpunkt steht insbesondere die Frage, welche Wirkung regionale Faktoren auf gesellschaftliche Differenzierungsprozesse haben.¹⁾ Jedoch fehlt für entsprechende empirische Analysen oftmals die geeignete Datenbasis. Der Mikrozensus erscheint daher mit seinem hohen Stichprobenumfang (1 % der Wohnbevölkerung) als eine gute Ergänzung zu den bereits vorhandenen Datenquellen der Regionalforschung.

In den vergangenen Jahren wurde der Mikrozensus in der Wissenschaft – wie auch der vorliegende Band dokumentiert – bereits für eine Reihe von Forschungsprojekten genutzt. Durch seinen inhaltlich breit gefächerten Merkmalskanon, insbesondere in den Bereichen Arbeitsmarkt, Bildung und Familie, bietet sich der Mikrozensus für unterschiedliche Forschungsprojekte im Bereich der Sozial- und Wirtschaftswissenschaften an. Zudem bietet er die Möglichkeit, Analysen auf Haushalts-, Familien- und Personenebene durchzuführen. Der Mikrozensus wird auf Grund seines Stichprobenumfangs von bundesweit ca. 800 000 Personen in ca. 350 000 Haushalten bereits vielfältig für die Untersuchung kleiner Subpopulationen, wie beispielsweise Ausländer unterschiedlicher Nationalität oder Personengruppen verschiedener Wirtschafts- oder Berufssparten, genutzt.

Neben diesen inhaltlich differenzierten Analysen bietet sich der Mikrozensus zudem für regional tief gegliederte Analysen an: Auch hier weist der Mikrozensus – im Verhältnis zu anderen sozialwissenschaftlichen Erhebungen – noch vergleichsweise hohe Fallzahlen auf.

Der folgende Beitrag zeigt das Potenzial des Mikrozensus für regional differenzierte Analysen auf und weist gleichzeitig auf einige methodische Aspekte des Mikrozensus hin, die beim Arbeiten auf Ebene kleiner räumlicher Einheiten beachtet werden müssen. Abschließend werden die vorhandenen und geplanten Zugangswege zu den Mikrodaten des Mikrozensus für regionalisierte Analysen dargestellt.

¹⁾ Vgl. zur Bedeutung der Regionalforschung beispielsweise Grözinger 2005.

2 Methodische Aspekte des Mikrozensus: Stichprobendesign und Stichprobenfehler

Eine wichtige Rolle bei der Bearbeitung von Fragestellungen, die sich auf kleinräumige Einheiten beziehen, spielt neben der Stichprobengröße das Stichprobendesign²⁾ der jeweiligen Erhebung. Beide haben Einfluss auf die Höhe des Stichprobenfehlers.

Bei den Mikrozensus-Erhebungen ab 1990 handelt es sich um einstufig geschichtete Klumpenstichproben. Die Klumpen bestehen aus Gebäuden oder Gebäudeteilen, den so genannten Auswahlbezirken. Es werden grundsätzlich vier Gebäudeschichten unterschieden: Schichten mit 1 bis 4, 5 bis 10 sowie 11 und mehr Wohnungen pro Gebäude.³⁾ Die vierte Gebäudeschicht umfasst Gemeinschaftsunterkünfte, wie z. B. Heime oder Anstalten. Im Durchschnitt besteht ein Auswahlbezirk aus neun Wohnungen. In den Wohnungen werden alle Haushalte und hier alle Haushaltsmitglieder erfasst. Als Grundlage für die Festlegung der Auswahlbezirke dienten die Gebäudedatei der Volkszählung 1987 für das alte Bundesgebiet und die Informationen aus dem Statistischen Register der DDR für die neuen Bundesländer. Weiterhin wird die Neubautätigkeit über die Bautätigkeitsstatistik mit zusätzlichen Auswahlbezirken berücksichtigt. Diese bilden eine weitere eigene Gebäudeschicht. Aus den so festgelegten Auswahlbezirken werden 1 % – Stichproben gezogen. Diese Stichproben sind wiederum in vier Teile gegliedert. Über ein Rotationsverfahren wird alle vier Jahre ein Viertel der Auswahlbezirke ausgetauscht. Das bedeutet, dass ein Haushalt, sofern er nicht aus bzw. in den Auswahlbezirk zieht, vier Jahre in der Mikrozensus-Erhebung enthalten ist.

Im Vergleich zu einer einfachen Zufallsauswahl von Personen oder Haushalten kann die Klumpenstichprobe zu einer Erhöhung des Zufallsfehlers führen.⁴⁾ Dies hat insbesondere Auswirkungen auf kleinräumige Analysen, da hier bereits mit hohen Zufallsfehlern auf Grund der geringen Fallzahlen pro regionaler Einheit zu rechnen ist.

In der amtlichen Statistik wurden auf Basis des Mikrozensus Untersuchungen zum Stichprobenfehler durchgeführt. Beispielrechnungen für NRW haben gezeigt, dass in einer Durchschnittsregion mit 500 000 Einwohnern der relative Standardfehler für den Merkmalsbereich Bevölkerung und Erwerbstätigkeit je nach Besetzungszahl erheblich ansteigt. Betrachtet man eine Subpopulation mit hochgerechnet 3 000 Fällen, beträgt der relative Standardfehler ca. 20 %, bei 10 000 immerhin noch 12 %. Bei 40 000 Fällen halbiert sich der relative Standardfehler auf 6 % und bei einer Subpopulation von 100 000 ergibt sich ein relativer Standardfehler von 4 %.⁵⁾ Von Seiten der amtlichen Statistik werden keine Ergebnisse mit einem relativen Standardfehler von über 15 % veröffentlicht. Aus diesem Grund beschränkt sich die amtliche Statistik bei der Berechnung regionaler Ergebnisse auf Basis des Mikrozensus zumeist auf Regionen mit mindestens 200 000 Einwohnern. Die entsprechenden regionalen Einheiten sind die so genannten Anpassungsschichten, auf deren Basis ebenfalls die obige Fehlerrechnung durchgeführt wurde. Diese Anpassungsschichten basieren auf Zusammenfassungen von kreisfreien Städten und Kreisen und umfassen durchschnittlich 500 000 Einwohner. Für diese Schichten werden im Mikrozensus entsprechende Hochrechnungsfaktoren bereitgestellt, die es erlauben, die Ergebnisse des Mikrozensus an die Zahlen der Bevölkerungsfortschreibung anzupassen.⁶⁾

2) Vgl. zum Auswahlplan des Mikrozensus ab 1990: Meyer 1994. – 3) Die Auswahlbezirke umfassen bei Gebäuden mit 11 und mehr Wohnungen nur einen Teil der Wohnungen des Gebäudes. – 4) Vgl. Meyer 1994 und Statistisches Bundesamt 2003. – 5) Vgl. Müller 1992 sowie für Baden-Württemberg Werner 1994. – 6) Vgl. zum Hochrechnungsverfahren im Mikrozensus Heidenreich 1994.

3 Regionale Gliederungen im Mikrozensus und ihr Analysepotenzial

Auch wenn man die im vorherigen Abschnitt dargestellten methodischen Aspekte beim Arbeiten mit dem Mikrozensus berücksichtigen sollte und sich hieraus entsprechende Einschränkungen bei der Interpretation der Ergebnisse ergeben können, bietet der Mikrozensus vielfältige Möglichkeiten für die Regionalforschung. Zum einen ist im Mikrozensus eine Vielzahl räumlicher Abgrenzungsmöglichkeiten auf tiefer regionaler Ebene enthalten, die es beispielsweise zulassen, spezifische Aggregationen für eigene Forschungsvorhaben vorzunehmen. Diese werden im Folgenden dargestellt. Zum anderen verfügt der Mikrozensus inhaltlich über ein breites Merkmalspektrum, das insbesondere für Analysen mit Bezug auf das Erwerbsleben geeignet ist. Exemplarisch wird hier eine Fragestellung aus dem Bereich der Erwerbstätigkeit aufgegriffen.

3.1 Regionale Gliederungen im Mikrozensus

Die im Mikrozensus standardmäßig vorhandenen regionalen Einheiten orientieren sich im Wesentlichen an den bestehenden administrativen Gliederungen. Darüber hinaus sind aber weitere räumliche Abgrenzungen und Kategorisierungen verfügbar. Im Folgenden werden administrative, mikrozensus-spezifische und nichtadministrative Raumeinheiten unterschieden.

Administrative Raumeinheiten	Mikrozensus-spezifische Raumeinheiten	Nichtadministrative Raumeinheiten
<ul style="list-style-type: none"> • Bundesländer (16) • Regierungsbezirke (26) • Kreise/kreisfreie Städte (440) • Gemeinden (12 431) 	<ul style="list-style-type: none"> • Regionale Anpassungsschichten (135): Ø 500 000 Einwohner • Regionale Schichten (228): 200 000 bzw. 250 000 Einwohner • Regionale Untergruppen: mind. 100 000 Einwohner • Gebäudeschichten • Auswahlbezirke 	<p><u>Kreisebene:</u></p> <ul style="list-style-type: none"> • Arbeitsmarktregionen (225/271) • Raumordnungsregionen (97) • Siedlungsstrukturelle Kreistypen <p><u>Gemeindeebene:</u></p> <ul style="list-style-type: none"> • Siedlungsstrukturelle Gemeindetypen • Arbeitsamtbezirke • Gemeindegrößenklasse

Administrative Raumeinheiten:

Im Mikrozensus sind neben den 16 Bundesländern ebenfalls die 440 Kreise des Bundesgebiets ausgewiesen. Sofern ein Bundesland über Regierungsbezirke verfügt, können auch diese über ein zusätzliches Merkmal abgegrenzt werden. Aufgeführt sind ferner die Gemeinden.

Für regionale Analysen bieten sich in Nordrhein-Westfalen (NRW) beispielsweise insgesamt fünf Regierungsbezirke und 54 Kreise an (s. Abb. 1). Betrachtet man für diese Gebiete die im Mikrozensus vorhandenen Fälle,⁷⁾ so kann man bei der Analyse auf Ebene der Regierungsbezirke auf Angaben von bis zu ca. 50 000 Personen aus ca. 20 000 Haushalten zurückgreifen. Die 54 Kreise unterscheiden sich stark in der Höhe der vorhandenen Fallzahlen. Im größten Kreis stehen Angaben von ca. 7 700 Personen in ca. 4 000 Haushalten und im kleinsten Kreis immerhin noch Angaben von ca. 1 000 Personen in etwa 450 Haushalten zur Verfügung.

⁷⁾ Diese und die nachfolgenden Fallzahlen beziehen sich auf den Mikrozensus 2000.

Einwohner. NRW ist in insgesamt 34 Anpassungsschichten aufgeteilt (s. Abb. 2). Die Einwohnerzahl der einzelnen Anpassungsschichten schwankt jedoch zum Teil erheblich. Dieses führt dazu, dass in der größten Anpassungsschicht, der kreisfreien Stadt Köln, ca. 7 700 Personen in ungefähr 4 000 Haushalten befragt wurden, während in kleineren Anpassungsschichten wie z. B. Bonn oder Gelsenkirchen lediglich um die 2 200 Personen in ca. 1 000 Haushalten in den Mikrozensus mit einbezogen wurden.

Abb. 2 Nordrhein-Westfalen:
Anpassungsschichten



Für NRW ist bereits mit den Anpassungsschichten eine gute regionale Differenzierung möglich: Aus 54 einzelnen kreisfreien Städten bzw. Kreisen werden 34 Anpassungsschichten gebildet. In anderen Bundesländern setzen sich die Anpassungsschichten aus mehreren Kreisen zusammen und geben ein nicht ganz so differenziertes Bild wieder. Beispielsweise stehen in Bayern mit seinen 96 kreisfreien Städten und Landkreisen nur insgesamt 22 Anpassungsschichten zur Verfügung.

Weiterhin werden im Mikrozensus regionale Schichten, die sich an Raumeinheiten mit Einwohnerzahlen von 200 000 – 250 000 orientieren, sowie regionale Untergruppen mit mind. 100 000 Einwohnern unterschieden. Bei diesen Abgrenzungen handelt es sich wiederum zum größten Teil um einzelne kreisfreie Städte bzw. Kreise. Größere kreisfreie Städte bzw. Kreise werden jedoch nochmals in kleinere Teilbereiche untergliedert. Als methodische Kennzeichen werden zusätzlich die Gebäudeschichten und die Auswahlbezirksnummern im Mikrozensus abgebildet.

Nichtadministrative Raumeinheiten:

Neben den bisher beschriebenen Raumeinheiten werden im Mikrozensus weitere Regionen und Raumkonzepte umgesetzt. Dazu zählen beispielsweise Arbeitsmarktregionen und Raumordnungsregionen. Arbeitsmarktregionen sind Zusammenfassungen von Stadt- und Landkreisen. Sie wurden über die Analyse von Pendlerbeziehungen zwischen Wohn- und Arbeitsort zunächst auf Gemeindeebene ermittelt und dann auf Kreisebene übertragen. Die Zuordnung zu einer Arbeitsmarktregion erfolgt unter Berücksichtigung des regionalen Arbeitsmarktzentrums und der Festlegung einer maximalen Pendlerentfernung. Insgesamt werden 225 bzw. 271 Arbeitsmarktregionen im Mikrozensus für das gesamte Bundesgebiet ausgewiesen.⁸⁾ Die 97 Raumordnungsregionen des Bundesgebietes sind wesentlich größere Gebietseinheiten, die eine Stufe zwischen den Regierungsbezirken und den Kreisen darstellen. Sie orientieren sich an den funktionalen Verflechtungen der Kreise sowie zusätzlich an der Aufteilung der Bundesländer in Planungsregionen. Darüber hinaus werden im Mikrozensus die siedlungsstrukturellen Gebietstypen ausgewiesen. Diese ermöglichen es, verschiedenartig strukturierte Siedlungsräume miteinander zu vergleichen. Die übergeordneten Gebietstypisierungen erfolgen auf der Basis der Verdichtung eines Gebietes, hier ist die Bevölkerungsdichte maßgeblich und die Zentralität, d. h. welche Stellung ein Gebiet für die umliegende Region hat. Es werden entsprechend Agglomerationsräume, verstärkte Räume und ländliche Räume unterschieden. Für die Typisierung eines Kreises ist neben diesen Kriterien weiterhin von Bedeutung, welchem Typ die dem Kreis übergeordnete Region angehört. Innerhalb der drei allgemeinen Gebietstypen werden für Kreise insgesamt neun Typen unterschieden. Im Bereich der Agglomerationsräume werden vier Ausprägungen unterschieden: Kernstädte, hochverdichtete, verdichtete und ländliche Kreise. Bei den ländlichen Räumen wird lediglich zwischen ländlichen Kreisen höherer und geringerer Dichte unterschieden. Eine analoge Typisierung mit insgesamt 17 Ausprägungen gibt es für die Gemeinden. Die siedlungsstrukturellen Gebietstypen und die Raumordnungsregionen wurden vom Bundesamt für Bauwesen und Raumordnung (BBR) entwickelt und finden in der Forschung vielfältige Anwendung.⁹⁾

Für die Regionalforschung ist zudem die Einteilung der Gemeinden in insgesamt neun unterschiedliche Gemeindegrößenklassen beginnend mit Gemeinden unter 2 000 Einwohnern und endend mit den Gemeinden über 500 000 Einwohnern von Interesse.

Auf der Grundlage der dargestellten regionalen Kategorien bietet sich im Mikrozensus zum einen die Möglichkeit, eigenständig unterschiedliche Regionen abzugrenzen. Zum anderen kann mit gängigen regionalen Typisierungen gearbeitet werden. In welcher Tiefe und Differenziertheit dieses bereits jetzt schon für einzelne Projekte möglich ist, wird im abschließenden Abschnitt erläutert.

8) Ab dem Mikrozensus 2003 wird die Einteilung in insgesamt 225 durch eine Unterteilung in 271 Arbeitsmarktregionen abgelöst. Für weitergehende Informationen zum Thema Arbeitsmarktregionen vgl. Eckey 1991. – 9) Nähere Informationen zu den Regionen und Raumkonzepten sind auf den Internetseiten des BBR zu erhalten: <http://www.bbr.bund.de/>.

3.2 Inhaltliche Analysen auf Basis des Mikrozensus

Die regionalen Gliederungen im Mikrozensus machen es möglich, unterschiedliche soziale Entwicklungen in einzelnen Regionen des Bundesgebietes zu betrachten. Ein inhaltlicher Schwerpunkt im Mikrozensus liegt auf den Fragen zur Erwerbsbeteiligung der Personen. Im Folgenden wird exemplarisch für einzelne Regionen in NRW die Erwerbstätigenquote betrachtet. Sie vermittelt einen ersten Eindruck über die Arbeitsmarktsituation in den betrachteten Gebieten. Analyseebene sind die Anpassungsschichten. Es wurden insgesamt vier Anpassungsschichten ausgewählt: Borken/Steinfurt, Coesfeld/Warendorf/Münster, Bottrop/Recklinghausen und Gelsenkirchen. Zusammengefasst bilden sie den Regierungsbezirk Münster. Die Erwerbstätigenquote beschreibt den Anteil der erwerbstätigen Personen an der Bevölkerung im erwerbsfähigen Alter. Als erwerbstätig werden im Mikrozensus all die Personen bezeichnet, die in der Berichtswoche mindestens eine Stunde gearbeitet haben – unabhängig ob als Selbstständiger oder Arbeitnehmer – und älter als 15 Jahre sind.¹⁰⁾

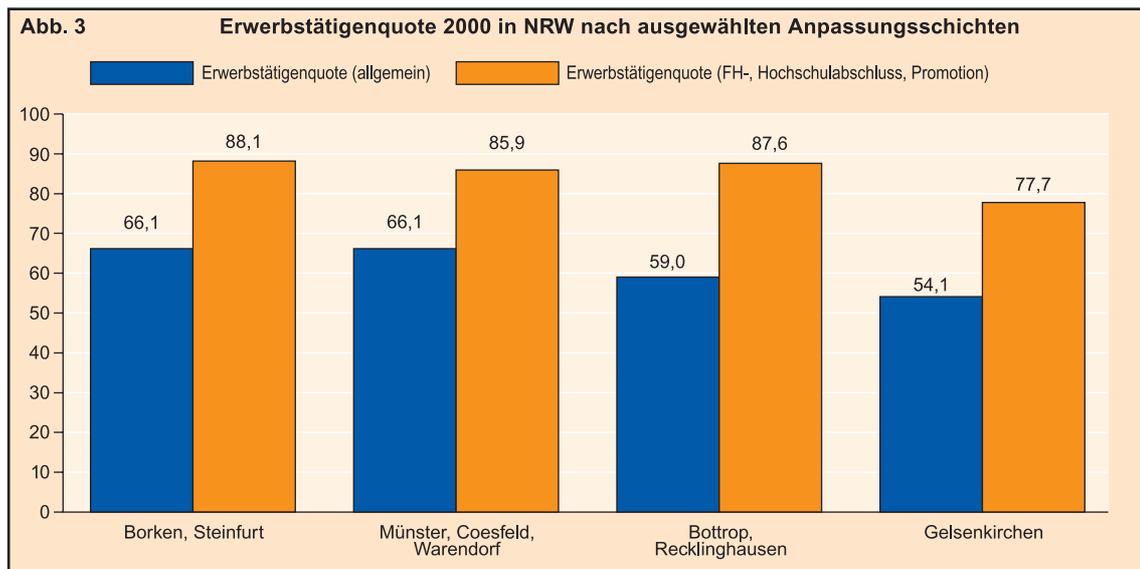
1. Erwerbstätigenquote 2000 in NRW nach ausgewählten Anpassungsschichten					
NRW	Reg.-Bez. Münster	Anpassungsschichten			
		Borken/Steinfurt	Coesfeld/Warendorf/Münster	Bottrop/Recklinghausen	Gelsenkirchen
62,9	62,7	66,1	66,1	59,0	54,1

Quelle: Mikrozensus 2000 (NRW), Personen am Ort der Hauptwohnung

Während die Erwerbstätigenquote in NRW und im Regierungsbezirk Münster mit 62,9 bzw. 62,7 noch recht ähnlich ist, erhält man bei der Betrachtung der einzelnen Anpassungsschichten ein wesentlich differenzierteres Bild. In den Anpassungsschichten Borken/Steinfurt sowie Coesfeld/Warendorf/Münster liegt die Erwerbstätigenquote mit 66,1 deutlich über dem Niveau von NRW und den beiden übrigen Gebieten des Regierungsbezirks Münster. In Bottrop/Recklinghausen sind 59 % der Personen im erwerbsfähigen Alter erwerbstätig; in Gelsenkirchen lediglich 54,1 % (s. Tab. 1). Dieses Ergebnis deutet bereits auf wirtschaftliche und strukturelle Unterschiede in den betrachteten Gebieten hin.

Geht man einen Schritt weiter und betrachtet die Erwerbstätigenquote im Hinblick auf den Bildungshintergrund der Erwerbstätigen, so ergibt sich ein noch differenzierteres Bild. Auf Grund von Erkenntnissen in der Arbeitsmarktforschung ist zu erwarten, dass sich die Arbeitsmarktsituation für beruflich höher qualifizierte Personen wesentlich besser darstellt als für andere Personengruppen. Es stellt sich aber die Frage, ob die positive Stellung dieser Personen auf dem Arbeitsmarkt in den Regionen unterschiedlich stark ausgeprägt ist. Betrachtet wird im Folgenden die Erwerbstätigenquote von Personen mit einem Fach-, Hochschulabschluss oder einer Promotion als höchsten beruflichen Abschluss im Vergleich zu der Erwerbstätigenquote aller Personen. Insgesamt zeigt sich für das Gebiet von NRW, dass die Erwerbstätigenquote für höher qualifizierte Personen mit 85,9 – wie zu erwarten – deutlich höher ist als die Quote aller Personen im erwerbsfähigen Alter. Der Unterschied beträgt ca. 23 Prozentpunkte und verdeutlicht eindrucksvoll die Besserstellung der höher qualifizierten Personen im Erwerbsleben. Das gleiche Verhältnis bildet sich für den Regierungsbezirk Münster ab. In der Abbildung 3 sind die Erwerbstätigenquoten der vier Anpassungsschichten im Vergleich abgebildet. Hier wird sichtbar, dass die Erwerbstätigenquoten der höher qualifizierten Personen in drei Anpassungsschichten gleich dem

10) Personen, die in der Berichtswoche aus verschiedenen Gründen wie z. B. Urlaub keiner Erwerbstätigkeit nachgegangen sind, sonst aber erwerbs- oder berufstätig sind, werden ebenfalls als Erwerbstätige betrachtet. Zur Erwerbstätigkeit im Mikrozensus vgl. Schmidt 2000.



Quelle: Mikrozensus 2000 (NRW), Personen am Ort der Hauptwohnung

Niveau von NRW sind oder sogar etwas darüber liegen. Regionale Unterschiede ergeben sich aber jeweils im Verhältnis zur allgemeinen Erwerbstätigenquote. In Borken/Steinfurt mit einer relativ hohen und in Gelsenkirchen mit einer eher niedrigen allgemeinen Erwerbstätigenquote ist das Verhältnis zwischen allen Erwerbstätigen und den höher qualifizierten Erwerbstätigen vergleichbar mit dem Ergebnis für gesamt NRW. Auch hier beläuft sich der Unterschied zwischen den beiden Quoten auf 22 bis 23,6 Prozentpunkte. In Bottrop/Recklinghausen stellt sich dieses Verhältnis anders dar: Mit einer Erwerbstätigenquote für Höherqualifizierte von 87,6 liegt dieser Wert 28,6 Prozentpunkte über der Quote aller Personen. Höher qualifizierte Personen sind im Vergleich zu allen anderen Personen 1,5 mal so häufig erwerbstätig. Eine niedrige allgemeine Erwerbstätigenquote geht demnach nicht notwendigerweise mit einer niedrigen Erwerbstätigenquote für höher qualifizierte Personen einher. In Bottrop/Gelsenkirchen finden diese bessere Arbeitsmöglichkeiten vor als es in den anderen Regionen der Fall ist. Allerdings stellt sich die Frage, ob die hohe Erwerbsbeteiligung der Höherqualifizierten zu Lasten aller anderen Personen geht und weniger qualifizierte Personen entsprechend seltener einen Arbeitsplatz finden. In der Anpassungsschicht Coesfeld/Warendorf/Münster liegen die Erwerbstätigenquoten aller Personen und die der höher qualifizierten am nächsten beieinander, was sich zum Teil dadurch erklärt, dass der Anteil der Höherqualifizierten an den Erwerbstätigen in dieser Anpassungsschicht fast doppelt so hoch ist als in den anderen Anpassungsschichten.

Allerdings muss bei dieser Betrachtung der Effekt des Stichprobenzufalls berücksichtigt werden: Für die kleinste Anpassungsschicht beträgt der relative Standardfehler für die Gruppe der Höherqualifizierten bereits ca. 12 %.

Dieses Beispiel zeigt, dass sich die Arbeitsmarktchancen höher qualifizierter Personen regional unterschiedlich darstellen. Welche Mechanismen in den einzelnen Regionen zum Tragen kommen, muss in weiteren Untersuchungen betrachtet werden. Die Analysen müssen dann beispielsweise die Zusammensetzung der Erwerbstätigen in den Regionen, die lokalen Arbeitsmarktbedingungen und die Verflechtungsbeziehungen zu anderen Regionen stärker mit einbeziehen sowie nach Art der Erwerbstätigkeit (z. B. Voll- oder Teilzeit) differenzieren.¹¹⁾

11) Vgl. zur Erwerbstätigkeit in NRW auf Anpassungsschichtebene Hullmann 1997.

4 Nutzungswege für die Wissenschaft

Der Mikrozensus steht Wissenschaftlerinnen und Wissenschaftlern als Datenquelle für Forschungsprojekte bereits seit einigen Jahren zur Verfügung. Er wird vor allem in Form des für mehrere Mikrozensususerhebungen vorhandenen Scientific-Use-File (SUF)¹²⁾ genutzt. Es handelt sich hierbei um faktisch anonyme¹³⁾ Mikrodatenfiles, die von den Wissenschaftlerinnen und Wissenschaftlern unter den geltenden rechtlichen Bestimmungen (Bundesstatistikgesetz § 16, Absatz 6) an ihrem Arbeitsplatz in einer wissenschaftlichen Einrichtung analysiert werden können. Diese so genannten Mikrozensus-Grundfiles zeichnen sich durch eine tiefe fachliche Gliederung aus. Es findet sich beispielsweise im Bereich der Berufe eine Differenzierung von bis zu 350 der 369 unterschiedlichen Berufsordnungen der Klassifizierungen der Berufe von 1992¹⁴⁾ und es werden 19 einzelne Staatsangehörigkeiten sowie weitere Zusammenfassungen von Nationalitäten unterschieden. Aus Gründen der faktischen Anonymität wird aber auf eine tiefe regionale Differenzierung verzichtet. Im Mikrozensus-Grundfile kann lediglich zwischen den einzelnen Bundesländern¹⁵⁾ unterschieden werden. Darüber hinaus enthält es die für Regionalforscher interessante Unterscheidung von fünf Gemeindegrößenklassen¹⁶⁾ sowie fünf Gebäudegrößenklassen¹⁷⁾. Ebenfalls steht die Auswahlbezirksnummer in systemfreier Form zur Verfügung. Diese ermöglichen eine sehr kleinräumige Betrachtung der einzelnen Klumpen des Mikrozensus. Die Auswahlbezirke sind über die Bundesländer hinaus jedoch regional nicht zuzuordnen.

Für Wissenschaftlerinnen und Wissenschaftler, die regional tiefer gehende Analysen mit dem Mikrozensus durchführen wollen, besteht seit geraumer Zeit die Möglichkeit über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder mit den Daten zu arbeiten. Die Mikrodaten können beispielsweise an einem Gastwissenschaftlerarbeitsplatz ausgewertet werden. Da die Daten in diesem Fall die Räume der amtlichen Statistik nicht verlassen und zusätzliche Vorkehrungen zur Sicherung der Daten vorgenommen werden, kann hier die faktische Anonymität mit weit geringerem Aufwand gewährleistet werden als bei den bestehenden SUF. Für jedes Forschungsprojekt wird ein spezifisches Datenfile erstellt, das lediglich die benötigten Variablen enthält. Dieses Vorgehen lässt es auch zu, regional tiefergehende Informationen zur Verfügung zu stellen. Gleiches gilt für die kontrollierte Datenfernverarbeitung: Hier besteht kein direkter Kontakt zwischen Wissenschaftlerin bzw. Wissenschaftler und Mikrodaten, sodass sich vielfältige Möglichkeiten für differenzierte Analysen anbieten. Die Ergebnisse werden nach Prüfung auf geheim zuhaltende Informationen an die Wissenschaftlerinnen und Wissenschaftler weitergeleitet.

Im Bezug auf regionale Analysen bieten sich über die Forschungsdatenzentren zurzeit Analysen auf Basis der Regierungsbezirke, der Raumordnungsregionen und als kleinste regionale Abgrenzung der Anpassungsschichten an. Es ist aber auch möglich, eigene Aggregationen, die sich an den administrativen Raumeinheiten orientieren, in den Daten umzusetzen und zu analy-

12) Als SUF stehen zurzeit die Mikrozensususerhebungen der Jahre 1989, 1991, 1993, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2003 und 2004 zur Verfügung. – 13) Faktisch anonym heißt, dass eine Reidentifikation von Personen im Mikrozensus nur mit einem unverhältnismäßig hohen Aufwand möglich ist. (Vgl. Müller 1991) – 14) Vgl. Statistisches Bundesamt 1992. – 15) Zusätzliche kann mit Hilfe einer weiteren Variablen zwischen Ost- und Westberlin unterschieden werden. – 16) Die Gemeindegrößenklasse enthält folgende Merkmalsausprägungen: weniger als 5 000, 5 000 bis 20 000, 20 000 bis 100 000, 100 000 bis 500 000 und mehr als 500 000 Einwohner. – 17) Die Gebäudegrößenklasse unterscheidet zwischen kleinen Gebäuden, mittleren Gebäuden und Großgebäuden sowie in Gemeinschafts- und Anstaltsunterkünften. Eine weitere Gebäudeklasse enthält die Neubauten.

sieren. Populationsschätzungen sind aber derzeit lediglich auf dem Niveau der Anpassungsschichten möglich, da nur für diese Schicht entsprechende Hochrechnungsfaktoren vorliegen.

Zurzeit führen das Zentrum für Umfragen, Methoden und Analysen (ZUMA) und das Forschungsdatenzentrum der Statistischen Landesämter ein Projekt zur Erstellung eines Mikrozensus-Regionalfiles als SUF durch. Mit dem SUF werden Analysen auf kleineren regionalen Einheiten wie den Anpassungsschichten möglich sein. Die tiefste regionale Gliederung des Mikrozensus-Regionalfiles sind die Mikrozensuskreisregionen (MZKR). Die MZKR wurden vom Bundesamt für Bauwesen und Raumordnung (BBR) in Kooperation mit ZUMA und den statistischen Ämtern entwickelt. Ausgangspunkt der regionalen Einheiten sind Kreise bzw. kreisfreie Städte. Die Kreise müssen laut Anonymisierungskonzept für das SUF¹⁸⁾ eine Mindestzahl von 100 000 Einwohnern aufweisen. Sofern die Kreise diese Einwohnerzahl nicht erreichen, werden sie mit anderen Kreisen zusammengefasst. Im Gegensatz zu den im Mikrozensus bereits enthaltenen regionalen Untergruppen, die ebenfalls mindestens 100 000 Einwohner umfassen, orientiert sich die Zusammenfassung der MZKR an den Raumordnungsregionen und den siedlungsstrukturellen Kreistypen. Insgesamt kann nach 349 MZKR differenziert werden.

Neben der für das Analysepotenzial möglichst schonenden Umsetzung des Anonymisierungskonzepts, das über die Vorgabe zur Größe der regionalen Einheiten hinaus insbesondere zu Zusammenfassungen im Bereich der Nationalität, der Berufe und der Wirtschaftszweige sowie des Alters führen wird, beschäftigt sich das Projekt mit einigen methodischen Fragestellungen. Zum einen wird ein Hochrechnungsfaktor für die MZKR entwickelt. Zum anderen wird betrachtet, in welcher regionalen Tiefe die Ergebnisse des Mikrozensus noch eine ausreichende Genauigkeit aufweisen. Geplant sind hier insbesondere Analysen zur Problematik des Stichprobenfehlers.

5 Ausblick

Mit den derzeitigen Arbeiten an der Erweiterung der Analysemöglichkeiten des Mikrozensus für Fragestellungen der Regionalforschung soll eine maßgebliche Verbesserung des bisherigen Datenangebots in diesem Bereich erreicht werden. Mit der Umsetzung eines Hochrechnungsverfahrens auf MZKR-Ebene wird die Möglichkeit gegeben, Populationsschätzungen auch für regionale Abgrenzungen durchzuführen, die sich nicht an den Abgrenzungen der Anpassungsschichten orientieren. Gleichwohl sind – wie erste Ergebnisse des Projekts bereits zeigen – der Populationsschätzungen auf Kreisebene Grenzen gesetzt. Der Mikrozensus-Regionalfile wird auf kleinräumiger Ebene deshalb vor allem Basis für die Analyse von Zusammenhangshypothesen sein können.

Ein erstes Mikrozensus-Regionalfile als SUF ist für die Erhebung aus dem Jahre 2000 vorgesehen. Dieses soll der wissenschaftlichen Forschung im Jahre 2006 zur Verfügung stehen. Für den Forschungsbedarf, der über die Möglichkeiten des Mikrozensus-Regionalfiles als SUF hinausgeht, stehen dann nach wie vor die weiteren Nutzungswege der Forschungsdatenzentren zur Verfügung.

18) Im Rahmen der Bestimmung des Anonymisierungskonzepts für das Mikrozensus-Grundfile als SUF wurden ebenfalls Empfehlungen für die Umsetzung eines Mikrozensus-Regionalfiles als SUF abgegeben. Vgl. Müller 1991. Vgl. zu diesem Thema ebenfalls Wirth 2004.

Literatur

Eckey, Hans-Friederich; Klemmer, Paul: Neuabgrenzung von Arbeitsmarktregionen für die Zwecke der regionalen Wirtschaftspolitik; in: Bundesamt für Bauwesen und Raumordnung: Informationen zur Raumentwicklung Heft 9/10; 1991, S. 569 – 577.

Grözinger, Gerd; Matiaske, Wenzel (Hg.) (2005): Deutschland regional. Sozialwissenschaftliche Daten im Forschungsverbund; München und Mering, Rainer Hampp Verlag.

Heidenreich, Hans-Joachim (1994): Hochrechnung des Mikrozensus ab 1990; in: Siegfried Gabler, Jürgen Hoffmeyer-Zlotnik u. Dagmar Krebs (Hg): Gewichtung in der Umfragepraxis, Opladen, Westdeutscher Verlag, S. 112 –123.

Hullmann, Alfred (1997): Privathaushalte und Erwerbstätigkeit in den Regionen Nordrhein-Westfalens, Regionalergebnisse des Mikrozensus; in: LDS NRW: Statistische Rundschau NRW; Ausgabe Juni 1997.

Meyer Kurt (1994): Zum Auswahlplan des Mikrozensus ab 1990; in: Siegfried Gabler, Jürgen Hoffmeyer-Zlotnik u. Dagmar Krebs (Hg): Gewichtung in der Umfragepraxis, Opladen, Westdeutscher Verlag, S. 112 – 123.

Müller, Alexander (1992): Regionalauswertungen aus dem Mikrozensus; in: Statistische Rundschau NRW, S. 179 – 189.

Müller, Walter; Blien, Uwe; Knoche, Peter; Wirth, Heike; u. a. (1991): Die faktische Anonymität von Mikrodaten; Forum der Bundesstatistik 19. Stuttgart: Metzler-Poeschel.

Schmidt, Simone (2000): Erwerbstätigkeit im Mikrozensus. Konzepte, Definitionen, Umsetzung; ZUMA-Arbeitsbericht 2000/01, Mannheim.

Statistisches Bundesamt (2003): Bevölkerung und Erwerbstätigkeit. Haushalte und Familien. Ergebnisse des Mikrozensus 2002. Fachserie 1. Reihe 3. Stuttgart. Metzler-Poeschel.

Statistisches Bundesamt (1992): Klassifizierung der Berufe. Systematisches und alphabetisches Verzeichnis der Berufsbenennungen, Stuttgart.

Werner, Joachim (1994): Regionalisierung des Mikrozensus; in: Baden-Württemberg in Wort und Zahl, S. 278 – 285.

Wirth, Heike; Zühlke, Sylvia; Christians, Helga (2004): Der Mikrozensus als Datenbasis für die Regionalforschung; in: Grözinger, Gerd; Matiaske, Wenzel (Hg.): Deutschland regional, Sozialwissenschaftliche Daten im Forschungsverbund; München und Mering, Rainer Hampp Verlag.

Autoren und Kontaktangaben:

Helga Christians, FDZ der Statistischen Landesämter, Landesamt für Datenverarbeitung und Statistik NRW, E-Mail: Helga.Christians@lds.nrw.de

Small Area Estimation beim Zensus 2011

Kersten Magg, Ralf Münnich, Josef Schäfer

Der Zensustest 2001, bei dem die Qualität der kommunalen Melderegister überprüft und Verfahren zur Durchführung eines künftigen registergestützten Zensus erprobt werden sollte, führte zu zwei unterschiedlichen Erkenntnissen. Die erprobten Verfahren, insbesondere die Durchführung einer Haushaltegenerierung aus einer Verknüpfung von Melderegistern und einer postalisch bei den Gebäudeeigentümern erhobenen Gebäude- und Wohnungszählung haben sich bewährt und können flächendeckend eingesetzt werden. Auf der anderen Seite erfordern die extremen Differenzen in der Qualität der Melderegister zwischen den Gemeinden und deren tendenzielle Abhängigkeit von der Gemeindegröße eine primärstatistische Ergänzung in Form einer Zufallsstichprobe. Diese Stichprobe soll gleichzeitig dazu genutzt werden, weitere, nicht in Registern enthaltene Merkmale, zu erheben. Zur Eingrenzung des benötigten Stichprobenumfangs und angesichts der in der Regel besseren Qualität der Melderegister in kleineren Gemeinden soll eine solche Erhebung im vollen Umfang allerdings nur in Gemeinden ab 10 000 Einwohnern erfolgen. In kleineren Gemeinden genügt zur Sicherstellung von Kreisergebnissen ein wesentlich geringerer Stichprobenumfang. Der Wunsch nach Ergebnissen in tiefer regionaler und sachlicher Gliederung kann dabei durch klassische Schätzverfahren allein, bei Wahrung hoher Qualitätsansprüche, häufig nicht befriedigt werden. Hier kämen Small Area-Verfahren zum Zuge, deren Einsatz im Rahmen eines registergestützten Zensus zu prüfen wäre und deren unterschiedliche Ansätze im Folgenden beschrieben werden.

Ausgangslage

Die vorerst letzte Volkszählung als Befragung aller Haushalte und Personen fand in West-Deutschland 1987, in Ost-Deutschland sogar bereits 1981 statt. Obwohl die Wiedervereinigung den Bedarf an vollständigen Strukturdaten noch erhöht hat und die Arbeitsgruppe *künftige Zensen* 1995 eine Kombination aus Vollerhebung und Stichprobe empfohlen hat, entschied sich die Bundesregierung 1996 gegen eine weitere traditionelle Volkszählung. Gründe dafür waren die hohen Kosten einer Totalerhebung sowie die Befürchtung einer fehlenden Akzeptanz in der Bevölkerung angesichts der damals noch nicht vergessenen Diskussion um die Volkszählung 1987. Stattdessen wurde die amtliche Statistik in Bund und Ländern aufgefordert, alternative Erhebungswege zu einer direkten Befragung aller Einwohner zu entwickeln. Der Empfehlung der Europäischen Union, im Jahr 2001 in allen Mitgliedsstaaten und Beitrittskandidatenländern einen flächendeckenden Zensus durchzuführen, ist Deutschland nicht gefolgt. Zunächst sollte im Rahmen einer Testerhebung auf Basis des Zensusvorbereitungsgesetzes¹⁾ die Qualität der Melderegister als Basis eines künftigen registergestützten Zensus geprüft und alternative Verfahren zu einer Totalerhebung aller Haushalte und Personen entwickelt und erprobt werden. Eine entsprechende Erhebung, der so genannte Zensustest, fand mit Stichtag 5. Dezember 2001 statt.

1) Gesetz zur Vorbereitung eines registergestützten Zensus vom 27. Juli 2001 (BGBl. I, S. 1882)

Die Überprüfung der Qualität und der Statistikauglichkeit der kommunalen Einwohnermelderegister, welche die Basis eines künftigen registergestützten Zensus darstellen, erfolgte mit Hilfe einer Zufallsstichprobe von bundesweit rund 38 000 Adressen in 555 Gemeinden. Für die ausgewählten Adressen wurden die Melderegisterauszüge mit den Feststellungen einer gleichzeitig durchgeführten traditionellen Haushaltebefragung verglichen und somit der Umfang an Karteileichen und Fehlbeständen in den kommunalen Melderegistern ermittelt.

Eine weitere Hauptaufgabe des Zensus-tests war die Entwicklung und Erprobung eines Modells, mit dem bei einem künftigen Zensus neben den auf Angaben zur Person beschränkten Registermerkmalen, auch Angaben zu Gebäuden, Wohnungen und Haushalten verfügbar gemacht werden sollen. Angaben zu Gebäuden und Wohnungen wurden primärstatistisch von den Gebäudeeigentümern oder -verwaltern mittels einer postalischen Gebäude- und Wohnungszählung (GWZ) erhoben. Um plausible Haushalte zu konstruieren, wurden die Angaben des Melderegisters adressenweise mit denen der GWZ zusammengeführt. Ferner wurden Angaben zur Erwerbstätigkeit aus den Registern der Bundesagentur für Arbeit herangezogen.

Dies erfolgte anhand einer Unterstichprobe von rund 16 000 Adressen in 222 Gemeinden. Die Ergebnisse der GWZ wurden mit denjenigen der für diese Unterstichprobe um Angaben zu Gebäude und Wohnungen sowie um Angaben zur Erwerbstätigkeit erweiterten Haushaltebefragung verglichen. Außerdem erfolgte anhand dieser Unterstichprobe eine Bewertung der Qualität der Register der Bundesagentur für Arbeit.

Kernstück des getesteten Zensusmodells war eine Haushaltegenerierung aus einer Zusammenführung von Angaben aus den Einwohnermelderegistern mit den Informationen aus der GWZ – vor allem zur Zahl der Wohnungen, zu den Inhabern der jeweiligen Wohnung sowie der Zahl der Personen je Wohnung. Den Wohnungsinhabern wurden schrittweise Ehegatten und Kinder (beides ist im Melderegister verzeichnet als so genannte Verzeigerung) sowie Personen gleichen Familien- oder Geburtsnamens, gleichen Einzugsdatums oder anderen Merkmalen, die ein Zusammenwohnen erwarten lassen, zugeordnet. Ging die Zuordnung nicht auf, war dies ggf. Anlass für gezielte Rückfragen.

Die postalische GWZ bei den Eigentümern kann nach den Testerfahrungen als akzeptiertes, geeignetes und effizientes Erhebungsverfahren angesehen werden. Die Erhebungsmerkmale für die Gebäude und Wohnungen sowie die für die Haushaltegenerierung notwendigen Hilfsmerkmale können mit hinreichender Qualität erhoben werden.

Die Zusammenführung von GWZ und Melderegister über Straße und Hausnummer konnte grundsätzlich erfolgreich getestet werden. Dies gilt auch für die Generierung von Wohnhaushalten, für die zusätzlich die Namen der Wohnungsinhaber einbezogen wurden. Durch die Einbeziehung der Angaben zu den Wohnungen konnten deutlich bessere Ergebnisse erzielt werden als bei den bisherigen von den Kommunen entwickelten, ausschließlich auf den Melderegistern beruhenden Generierungsverfahren. Verbliebene Ungenauigkeiten bei der Generierung von Wohnhaushalten, wie die Überschätzung großer Haushalte, resultierten im Wesentlichen aus den Fehlern (Karteileichen) der Melderegister.

Insgesamt lässt sich damit konstatieren, dass ein registergestützter Zensus nach dem getesteten Modell durchführbar ist. Der Verfahrenstest ist insofern als erfolgreich zu bewerten.

Qualität der Einwohnermelderegister (Registertest)

Als Ergebnis des so genannten Registertests, d. h. des Vergleichs der Melderegisterauszüge mit den Ergebnissen der Haushaltebefragung, wurden bundesweit Übererfassungen der Einwohnermelderegister (Karteileichen) im Umfang von 4,1 % sowie Untererfassungen (Fehlbestände) im Umfang von 1,7 % ermittelt, wobei die Ergebnisse allein schon zwischen Bundesländern und Gemeindegrößenklassen erheblich streuen (vgl. Tabelle 1). Die Streuung zwischen den Gemeinden dürfte noch wesentlich größer sein.

1. Karteileichen- und Fehlbestandsraten nach Bundesländern und Gemeindegrößenklassen					
Bundesland Gemeindegrößenklasse	Personen im Melderegister ¹⁾				
	insgesamt	Fehlbestände		Karteileichen	
	1 000		%	1 000	%
Baden-Württemberg	10 307,1	102,5	1,0	313,3	3,0
Bayern	11 957,5	211,6	1,8	418,1	3,5
Berlin	3 272,3	88,8	2,7	265,1	8,1
Brandenburg	2 542,4	25,1	1,0	94,2	3,7
Bremen	648,2	12,9	2,0	36,5	5,6
Hamburg	1 629,4	47,1	2,9	115,1	7,1
Hessen	5 801,2	138,9	2,4	268,4	4,6
Mecklenburg-Vorpommern	1 742,1	36,2	2,1	70,1	4,0
Niedersachsen	7 772,0	81,5	1,1	259,9	3,3
Nordrhein-Westfalen	17 408,8	350,3	2,0	754,4	4,3
Rheinland-Pfalz	3 972,3	50,3	1,3	126,3	3,2
Saarland	1 050,9	26,2	2,5	82,6	7,9
Sachsen	4 299,6	44,1	1,0	111,0	2,6
Sachsen-Anhalt	2 510,4	38,7	1,5	92,6	3,7
Schleswig-Holstein	2 724,1	84,8	3,1	169,8	6,2
Thüringen	2 346,4	29,4	1,3	60,2	2,6
Deutschland	79 984,9	1 368,4	1,7	3 237,5	4,1
davon					
Gemeinden mit ... bis unter ... Einwohnern					
unter 10 000	22 947,5	303,6	1,3	634,6	2,8
10 000 – 50 000	26 112,7	348,4	1,3	900,0	3,5
50 000 – 800 000	23 944,5	509,3	2,1	1 175,7	4,9
800 000 und mehr	6 980,2	207,1	3,0	527,2	7,6

1) Bevölkerung am Ort der Hauptwohnung; hochgerechnetes Ergebnis des Zensusstests vom 5. 12. 2001 – ohne Bevölkerung in Anstalten

Die festgestellte Karteileichenrate von 4,1 % konnte allerdings bereits im Rahmen eines registergestützten Zensus selbst durch folgende Maßnahmen erheblich reduziert werden:

- Aufdeckung der kurzfristigen umzugsbedingten Karteileichen durch einen Abgleich mit einem zweiten, zu einem späteren Zeitpunkt (im Test war dies knapp vier Monate nach dem Stichtag) erstellten Melderegisterauszug,
- Aufdeckung von mehrfach gemeldeten Personen durch eine bundesweite Mehrfachfallprüfung, d. h. eine Überprüfung auf Personen, die in mehr als einer Gemeinde mit Hauptwohnung gemeldet sind – deren Anzahl wurde im Rahmen des Zensusstests anhand einer separat durchgeführten Geburtstagsauswahl aus den Melderegistern aller deutscher Gemeinden ermittelt.
- Eine Nachfrage bei den Bewohnern von Adressen, in denen die Haushaltegenerierung keine plausiblen Wohnungszuordnungen vornehmen oder Haushaltszusammenhänge schaffen konnte.

Die genannten Maßnahmen senkten die Karteileichenrate im bundesweiten Durchschnitt von 4,1 % auf 1,8 %, sodass insgesamt nur noch ein geringer Unterschied zwischen den Karteilei-

chen- (1,8 %) und den Fehlbestandsraten (1,7 %) verbleibt. Allerdings bestehen nach wie vor erhebliche Differenzen zwischen den Gemeindegrößenklassen. Die Unterschiede zwischen einzelnen Gemeinden dürften noch wesentlich höher ausfallen.

2. Karteileichen- und Fehlbestandsraten nach Gemeindegrößenklassen und ersten Bereinigungsverfahren						
Gemeinden mit ... bis unter ... Einwohnern	Karteileichenrate ohne „temporäre“ Karteileichen	Zweite Bereinigung der Karteileichen durch			Karteileichenrate nach zweiter Bereinigung	Fehlbestandsrate
		Mehrfachfallprüfung	Haushaltgenerierung	insgesamt ¹⁾		
	%	Prozentpunkte			%	
unter 10 000	2,0	0,7	0,9	1,3	0,7	1,3
10 000 – 50 000	2,5	0,6	0,7	1,1	1,4	1,3
50 000 – 100 000	2,4	0,7	0,3	0,9	1,5	2,1
100 000 und mehr	4,3	0,6	0,4	0,9	3,4	2,4
Deutschland	2,9	0,6	0,6	1,1	1,8	1,7

1) Berechnet unter Verwendung eines Schätzwertes für die kombinierte Wirkung der Bereinigungsverfahren Mehrfachfallprüfung und Haushaltgenerierungen.

Die durchschnittliche Karteileichenrate in Gemeinden unter 10 000 Einwohnern betrug mit 0,7 % nur etwa ein Fünftel derjenigen in Gemeinden über 100 000 Einwohnern mit 3,4 %. Da die Fehlbestandsraten weniger deutlich zwischen den Gemeindegrößenklassen variieren als die Karteileichenraten, ergab sich im Durchschnitt ein Fehlbestandsüberhang in kleinen Gemeinden sowie ein Karteileichenüberhang in Großstädten. Ein Verzicht auf weitere Maßnahmen zur Ergänzung des künftigen Zensus würde für die kleinen Gemeinden tendenziell zu einer Unterschätzung der Einwohnerzahl, für die Großstädte dagegen zu deren Überschätzung führen.

Allein zur Reduzierung der Streuung von Karteileichen und Fehlbeständen zwischen den Gemeinden bedarf ein registergestützter Zensus daher zwingend primärstatistischer Ergänzungen – zumal eine zunächst erprobte Vorabausfilterung kritischer Adressen mit Hilfe multivariater Verfahren sich als wenig effektiv erwiesen hat.

Stichprobe in Gemeinden mit mindestens 10 000 Einwohnern

Zur Eingrenzung der Registerfehler auf Gemeindeebene bietet sich die Durchführung primärstatistischer Zusatzerhebungen als Zufallsstichproben von Registeradressen an. Ein konkretes Stichprobenmodell wurde im Ergebnisbericht zum Zensusstest bereits vorgeschlagen, kann aber aufgrund weitergehender Überlegungen, insbesondere auch im Hinblick auf den Einsatz von Small Area-Verfahren noch modifiziert werden. Gemäß diesem Vorschlag sollen je Gemeinde 550 Adressen zufällig aus den kommunalen Melderegistern ausgewählt werden.

Kosten-Nutzen-Überlegungen führten zur Empfehlung, diese Erhebungen auf Gemeinden mit mindestens 10 000 Einwohnern zu beschränken. Eine Durchführung der Stichprobe in allen Gemeinden entspräche einer Erhebung von gut 20 Millionen Personen, also rund einem Viertel der in Deutschland lebenden Menschen. Insbesondere in kleinen Gemeinden wären die Auswahltätze extrem hoch – bis hin zur Totalerhebung. Zudem sind die verbleibenden Registerfehler in kleinen Gemeinden deutlich geringer als in den Großstädten.

Erfolgt die Erhebung von jeweils 550 Adressen dagegen nur in den 1 513 Gemeinden mit mindestens 10 000 Einwohnern (Stand: 31. 12. 2000) sowie auf Kreisebene, lässt sich der Erhebungsumfang bei Einbeziehung weiterer Merkmale (550 Adressen je Gemeinde) auf rund 5,9 Mill. Personen verringern. Für die kleinen Gemeinden liegen dann allerdings weder Schätzungen der Karteileichen- und Fehlbestandsraten noch Informationen über weitere Merkmale vor.

Auswahleinheit der Stichprobe ist die Adresse im Melderegister der jeweiligen Gemeinde. Vorgeschlagen wurde – nach Proberechnungen mit unterschiedlichen Schichtungsmerkmalen – eine Schichtung nach der Zahl unterschiedlicher Familiennamen im Melderegister (die Haushaltszugehörigkeit ist im Melderegister unmittelbar nicht zu erkennen) sowie nach dem Zeitpunkt der letzten erfolgten Anmeldung vor Stichtag als Fluktuationsindikator. Diese führte zum Vorschlag folgender Schichtenbildung:

Schicht 1:	1 oder 2 Namen	letzte Anmeldung liegt mehr als 1 Jahr zurück
Schicht 2:	1 oder 2 Namen	letzte Anmeldung liegt höchstens 1 Jahr zurück
Schicht 3:	3 – 12 Namen	letzte Anmeldung liegt mehr als 1 Jahr zurück
Schicht 4:	3 – 12 Namen	letzte Anmeldung liegt höchstens 1 Jahr zurück
Schicht 5:	13 oder mehr Namen	

Aufgrund der geringen Fallzahl von Adressen mit 13 und mehr Namen wurde hierbei auf eine weitere Differenzierung verzichtet.

Die Auswahl der Schichtungsmerkmale sowie die Optimierung der Aufteilung der Gesamtstichprobe auf die Schichten (nach Neyman-Tschuprov) würde dann aufgrund der zentralen Bedeutung der Schätzung der amtlichen Einwohnerzahl anhand des Merkmals Saldo von Karteileichen und Fehlbeständen bei der Bevölkerung am Ort der Hauptwohnung erfolgen – ein Merkmal, das nach den Erkenntnissen des Zensusstests auf Adressebene eine höhere Streuung aufweist als die Merkmale Anzahl der Karteileichen und Anzahl der Fehlbestände. Karteileichen und Fehlbestände saldieren sich auf Adressebene demnach in der Regel nicht auf.

Die Entscheidung für die Zahl von 550 Adressen je Gemeinde resultierte dabei aus Schätzungen für einen Stichprobenumfang, der ausreicht, die Ausprägungen weiterer Merkmale, z. B. die Zahl der Selbstständigen, mit einer akzeptablen Genauigkeit als Anteilsschätzung auf Gemeindeebene zu ermitteln. Das primäre Kriterium der Ermittlung der amtlichen Einwohnerzahl mit einem relativen Standardfehler von $\pm 0,5\%$ im Durchschnitt der Gemeinden, wird in jeder Gemeindegrößenklasse mehr als hinreichend erfüllt. Der endgültige Stichprobenplan ist, wie bereits gesagt, noch nicht festgelegt. Der Umfang dieser Primärerhebung sollte sich jedoch am derzeit vorgeschlagenen Modell orientieren, da eine Kostenschätzung für die Ergänzungsstichprobe bereits in den Entscheidungsprozess für das weitere Vorgehen eingeflossen ist. Unter dieser Rahmenbedingung ist eine kostenoptimale Allokation der Stichprobe anzustreben. Die Planung des endgültigen Stichprobenmodells sollte dabei auch die Anforderung von Small Area-Verfahren berücksichtigen, sowohl zur Schätzung fachlich gegliederter Ergebnisse in großen Gemeinden als auch bezüglich eines möglichen Nachweises von Ergebnissen in Gemeinden unter 10 000 Einwohnern durch ein Herunterbrechen von Kreisergebnissen (s. u.) – allerdings ohne die konkurrierende Anforderung an eine möglichst gute Schätzung der amtlichen Einwohnerzahl zu vernachlässigen.

Ergebnisse für Gemeinden unter 10 000 Einwohnern

Zur Auswertung weiterer Merkmale auf Kreis-, Regierungsbezirks-, Landes- und Bundesebene ist grundsätzlich auch eine Erhebung in Gemeinden unter 10 000 Einwohnern erforderlich, da dieses Segment ansonsten unberücksichtigt bliebe. In jedem Kreis, der Gemeinden unter 10 000 Einwohnern enthält – und das sind fast alle – wäre eine Sonderschicht aus diesen Gemeinden zu bilden. Der Stichprobenumfang in den jeweiligen Gemeinden unter 10 000 Einwohnern entspricht dann dem 550-fachen Anteil an der Einwohnerzahl des Kreises. Bei Kreisen mit vielen kleinen und einzelnen großen Gemeinden kann dieser Wert ggf. sehr klein sein.

Diese kleine Zusatzerhebung allein reicht in den Gemeinden mit weniger als 10 000 Einwohnern zur hinreichend genauen Schätzung von Karteileichen und Fehlbeständen auf Gemeindeebene nicht aus. Die Ermittlung der Einwohnerzahlen würde in diesen Gemeinden dann nur auf Basis der Melderegisterauszählung erfolgen – ergänzt um die Erkenntnisse aus der Mehrfachfallprüfung, der Haushaltegenerierung sowie der GWZ. Da diese Bereinigungsverfahren in erster Linie die Zahl der Karteileichen verringern, sind noch Verfahren zum Erkennen von potenziellen Fehlbeständen in den kleinen Gemeinden und damit zur Verringerung des Fehlbestandsüberhangs zu entwickeln und zu überprüfen.

Angaben zu weiteren Merkmalen könnten ggf. aus den Kreisergebnissen oder den Ergebnissen anderer strukturell vergleichbarer Gemeinden heruntergebrochen werden. Auch wenn die Zahl der Auswahleinheiten in der jeweiligen Gemeinde beim derzeit vorgeschlagenen Stichprobenmodell äußerst klein ist, sollte die Frage nach möglichen Small Area-Verfahren gestellt werden, die geeignet sind, auch auf der Ebene von Gemeinden unter 10 000 Einwohnern hinreichend genaue Schätzungen für die Ausprägungen weiterer Merkmale bereitzustellen – auch wenn dies eine Modifikation des Stichprobenplans erfordert.

Die Anwendung klassischer Schätzverfahren stößt beim vorgeschlagenen Modell bei der Schätzung von Ergebnissen für Gemeinden unter 10 000 Einwohnern an Grenzen. Dieses Modell sieht die Erhebung von 550 Adressen auf der Ebene von Gemeinden ab 10 000 Einwohnern und einer minimalen Repräsentanz kleinerer Gemeinden, die lediglich die Ermittlung von Kreisergebnissen in einer der Ergebnisse größerer Gemeinden vergleichbaren Genauigkeit sicherstellt, vor. Ebenso ist zu befürchten, dass eine Anwendung klassischer Schätzmethoden der freien und teils auch der gebundenen Hochrechnung auf Grund zu erwartender hoher Standardfehler bei einer differenzierten sachlichen Untergliederung auch auf der Ebene von Gemeinden knapp über 10 000 Einwohnern kaum geeignete Ergebnisse erwarten lässt. Hinzu kommt, dass beim registergestützten Zensus die mit Fehlern behafteten ausgezählten Bevölkerungsregister mit Hilfe zusätzlicher Stichproben statistisch korrigiert werden. Bei einem klassischen Zensus, der auf einer Totalerhebung basiert, wäre dies nicht notwendig. Dabei müssen Karteileichen und Fehlbestände geeignet in eine Schätzung zur Korrektur der Registerdaten eingebunden werden. Daher wird der Einsatz spezieller Methoden, die für den Einsatz bei kleinen Nachweisgruppen besonders geeignet sind, diskutiert, die so genannten Small Area-Schätzmethoden.

Small Area Estimation

Wie der Begriff schon erkennen lässt, handelt es sich bei den interessierenden Größen um Parameter kleiner Subpopulationen. Diese Populationen können demographisch, aber auch geographisch abgegrenzt sein. Der Begriff der Small Area-Schätzung wird einerseits bei Schätzungen inhaltlich abgegrenzter kleiner Subpopulationen (soziodemographische Gruppe), in der Regel als Small Domain bezeichnet, sowie andererseits bei geographisch abgegrenzten kleinen Subpopulationen, den eigentlichen Small Areas, verwendet. Im Rahmen des Zensus spielen insbesondere Alter, Geschlecht und Nationalität als sachliche Untergliederung sowie Bundesländer, Regierungsbezirke, Kreise und Gemeinden in räumlicher Untergliederung eine bedeutende Rolle. Der konventionelle Sprachgebrauch sieht allerdings kaum eine getrennte Betrachtung vor, weswegen beide Methoden unter dem Begriff der Small Area-Schätzung subsumiert werden. Im Folgenden wird oftmals im Kontext der Small Areas oder Small Domains der Einfachheit halber von Schichten gesprochen.

Eine klare Abgrenzung, ab wann eine Small Area tatsächlich als klein erachtet werden kann, ist nicht eindeutig geklärt. Im Allgemeinen macht man es davon abhängig, ob es möglich ist, bei gegebenem Stichprobenumfang mit direkten Schätzverfahren zuverlässige Ergebnisse zu erzielen. Sollte dies nicht möglich sein, bedarf es der Anwendung spezieller Small Area-Methoden, weswegen man dann in diesem Fall von kleinen Subpopulationen (Small Domains oder Small Areas) spricht. Ebenso muss die Erklärung einer kleinen Subpopulation in Relation zur übergeordneten geographischen Region oder demographischen Gruppe gesehen werden (vgl. Fay/Herriot, 1979, S. 269 ff. und Münnich/Schmidt, 2002, S. 140). Ist eine Stichprobenerhebung auf eine große Population angepasst, kann der Fall eintreten, dass nur wenige oder keine Elemente der interessierenden Teilmenge in der Stichprobe enthalten sind. Im Kontext eines registergestützten Zensus in Deutschland könnte die Small Area-Schätzung bei der Schätzung einzelner Parameter kleiner Subpopulationen sowohl demographischer als auch geographischer Natur gewinnbringend eingesetzt werden – dies gilt ebenso für die Registerkorrektur in Bezug auf Karteileichen und Fehlbestände. Es kann in Erwägung gezogen werden, dass Small Area-Methoden verbesserte Schätzergebnisse allgemein auf Gemeindeebene, aber auch betreffend einzelner demographischer Subpopulationen oder Gemeindeteile, wie sie für größere Städte zur Diskussion stehen, liefern könnten.

Bisherige Anwendungen der Small Area-Schätzung

Für statistische Auswertungen hat die Small Area-Fragestellung schon seit Jahrzehnten mit wachsendem Gewicht große Bedeutung gewonnen. Bereits im 11. Jahrhundert soll es in England sowie im 17. Jahrhundert in Kanada erste derartige statistische Auswertungen gegeben haben (vgl. Ghosh/Rao, 1994, S. 55 und die darin zitierte Literatur), wenngleich die damalige Ausgangssituation und Zielsetzung aus heutiger Sicht von geringerer Bedeutung sind. Deutlich zunehmende wissenschaftliche Aktivitäten im Rahmen der Small Area-Forschung, vor allem im aktuellen anwendungsbezogenen Kontext, sind seit den 1960er-Jahren zu verzeichnen.

Die Einsatzgebiete dabei sind vielseitig: Beginnend bei Small Area-Statistiken, etwa im Hinblick auf Ausgleichszahlungen zwischen Kommunen und Land beziehungsweise Staat bis hin zu zuverlässigen Schätzungen von Konsumgewohnheiten ausgewählter kleiner soziodemographischer Gruppen. Länder wie beispielsweise die USA, Kanada oder Israel arbeiten schon intensiv

am Einsatz solcher Schätzmethoden. In den USA etwa kommen Small Area-Methoden bei verschiedenen Bundesprogrammen zum Einsatz: Infant and maternal health for states (NCHS), Personal income for states and counties (BEA), Post-census populations for counties (USCB) und viele andere (vgl. Lahiri, 2005, und Rao, 2003). Ein beliebtes Anwendungsgebiet von Small Area-Verfahren sind unter anderem die Schätzung von Bevölkerungszahlen als Grundlage zur Ermittlung staatlicher Transferzahlungen. In diesem Zusammenhang wurde 1980 das U.S. Bureau of the Census von mehreren Staaten und Städten der USA wegen zu gering ausgewiesenen Bevölkerungszahlen, insbesondere von Minderheiten, verklagt. Die Statistiker Ericksen, Kadane und Tukey, die auch als Gutachter vor Gericht auftraten, entwickelten verbesserte Verfahren, die zuverlässige Werte für das Ausmaß der Unterschätzung ethnischer Gruppen auf regionaler Ebene lieferten. Andere konkrete Einsatzgebiete sind etwa die Ermittlung verlässlicher Daten über die Verbreitung von Drogen- und Alkoholmissbrauch in ausgewählten demographischen Gruppen oder die vom U.S. Bureau of the Census im 2-Jahres-Rhythmus durchgeführte Schätzung von Armutsdaten, auf deren Grundlage beispielsweise über die Verteilung finanzieller staatlicher Mittel entschieden wird (vgl. Münnich/Schmidt, 2002, S. 139 und die darin zitierte Literatur). Speziell in der Arbeit von Schaible (1996) können weitere detaillierte Informationen über den Einsatz indirekter Schätzmethoden in den USA nachgelesen werden.

In Europa befassten sich zwei international ausgerichtete Forschungsprojekte, finanziert von der Europäischen Kommission im Rahmen des 5. Forschungsrahmenprogramms, ausführlich mit der Theorie sowie mit Einsatzmöglichkeiten, begleitet durch vielseitige Simulationsrechnungen und -auswertungen. Bei diesen Projekten handelt es sich einerseits um EURAREA (Enhancing Small Area Estimation Techniques to Meet European Needs; http://www.statistics.gov.uk/methods_quality/eurarea/), koordiniert vom Office for National Statistics und andererseits um das Projekt DACSEIS (Data Quality in Complex Surveys within the New European Information Society; <http://www.dacseis.de>), koordiniert vom Lehrstuhl für Statistik, Ökonometrie und Unternehmensforschung der Wirtschaftswissenschaftlichen Fakultät an der Eberhard-Karls Universität Tübingen. Im Projekt EURAREA beschäftigte man sich sehr intensiv mit der Methodik einzelner Small Area-Verfahren und untersuchte deren Anwendung am Beispiel zahlreicher Simulationsrechnungen. DACSEIS, das sich im Wesentlichen auf die Untersuchung der Datenqualität in komplexen Stichprobenerhebungen konzentrierte, untersuchte ebenfalls anhand umfangreicher Simulationsstudien die Einsatzmöglichkeiten der von EURAREA betrachteten Standardmethoden am Beispiel des deutschen Mikrozensus. Zusätzlich wurde der Einfluss von Antwortausfällen auf die Modellierungen evaluiert. Ausführliche Berichte über die Ergebnisse beider Forschungsprojekte können den jeweiligen Internetseiten entnommen werden.

Im Rahmen von Zensen wird zunehmend auch der Einsatz von Small Area-Verfahren diskutiert, wie etwa in der Schweiz (vgl. Renaud, 2004) und nun auch in Deutschland im Zusammenhang einer registergestützten Volkszählung.

Methodik der Small Area-Schätzung

Small Area-Schätzverfahren werden grundsätzlich in so genannte direkte und indirekte Methoden unterteilt. Schätzer, welche ausschließlich auf schichtspezifischen Stichprobendaten beruhen – d. h. auf Daten aus der jeweiligen interessierenden Small Area oder Small Domain – zählen zu den direkten Schätzern. Werden zusätzliche Informationen aus anderen Areas oder Domains, beispielsweise aus übergeordneten Populationen oder benachbarten bzw. ähnlichen Regionen oder einer anderen Zeitperiode zur Schätzung des interessierenden Parameters he-

rangezogen, dann spricht man von indirekten Schätzern (vgl. zu Small Area-Schätzung mit Zeitreihen im Speziellen Pfeffermann, 2005). Allein die Tatsache, dass Hilfsvariablen zur Schätzung herangezogen werden, ist nicht charakteristisch für die Zuordnung zu einer der genannten beiden Klassen.

Neben der Klassifizierung der Schätzverfahren in direkte und indirekte kennt man parallel dazu eine Einteilung in so genannte designbasierte, modellunterstützte sowie modellbasierte Schätzer. Diese beiden Charakterisierungen können sich teilweise überschneiden. Während sich die Unterteilung in direkte und indirekte Verfahren lediglich auf den Charakter der hinzugezogenen Informationen beschränkt, steht bei der zweiten Einteilung die Schätzmethodik im Vordergrund. Designbasierte Verfahren ergeben sich aus dem Design der Stichprobenerhebung und kommen ohne Verteilungsannahmen aus. Ein wesentlicher Vorteil dieser Schätzmethodik besteht in der zumeist vorhandenen Unverzerrtheit der Schätzer unter Berücksichtigung des Stichprobendesigns (vgl. Rao, 2003, S. 10). Man spricht dann von Design-Unverzerrtheit. Im Rahmen der modellbasierten Verfahren geht man bei endlicher Gesamtheit davon aus, dass sie Teil einer unendlichen Superpopulation ist und gewissen Verteilungsannahmen gehorcht. Es wird also der Schätzung bestimmter Parameter unterstellt, dass die abhängige Variable einer gewissen Verteilung folgt und mittels eines bestimmten Modells wiedergegeben werden kann. Dieses Modell beschreibt die Beziehung zwischen der Zielvariablen und den Hilfsvariablen über die gesamte Superpopulation. Schätzungen für kleine Schichten können direkt aus dem Modell der Superpopulation abgeleitet werden und können somit auch in Fällen generiert werden, in denen nur geringe, im Extremfall sogar keine, Stichprobeninformationen zur interessierenden Variablen vorliegen. Daraus kann allerdings auch ein wesentlicher Nachteil direkt abgeleitet werden. Modellbasierte Schätzungen sind in der Regel design-verzerrt und hängen sehr von der Validität der Modelle ab. Die dritte Kategorie von Schätzverfahren verwendet zur designbasierten Hochrechnung zusätzlich Modellvorstellungen. Der klassische Regressionsschätzer zählt zum Beispiel zur Klasse der modellunterstützten Schätzer, da er sich einerseits aus dem klassischen Horvitz-Thompson-Schätzer und andererseits aus einem Korrekturterm, resultierend aus einer linearen Einfachregression, zusammensetzt.

Unter der beschriebenen Klassifikation der Small Area-Schätzverfahren verbergen sich nun zahlreiche Methoden, die im Laufe der letzten Jahrzehnte entwickelt, angepasst und optimiert wurden. Die Literatur kennt dabei vor allem die expliziten direkten Methoden, die so genannten traditionellen demographischen Methoden und als explizit indirekte Verfahren synthetische Methoden, zusammengesetzte Schätzungen sowie die ebenfalls bereits genannten speziellen Small Area-Modellverfahren.

Demographische Methoden verwenden in der Regel aktuelle Daten aus administrativen Erhebungen in Verbindung mit Daten aus dem letzten Zensus. Die Theorie beschreibt nun mehrere Möglichkeiten, wie diese Daten miteinander verknüpft werden können. Stichworte wie Symptomatic Accounting Techniques, Regression Symptomatic Procedures oder Dual System Estimation of Total Population sollen an dieser Stelle genügen. Wenngleich diese Methoden per Definition nicht zu den Small Area-Modellen gezählt werden können, finden sie durchaus Anwendung in der Praxis der Small Area-Schätzung (vgl. Rao, 2003, S. 27 ff. und die darin zitierte Literatur). Mit der *Dual System-Schätzung* versuchen unter anderem beispielsweise die Statistiker in Israel zuverlässige Schätzergebnisse aufgrund des letzten vollständigen Zensus und zusätzlichen aktuellen Stichprobenerhebungen zu generieren (vgl. Nirel et al., 2003).

Ein synthetischer Schätzer charakterisiert sich grundsätzlich dadurch, dass ein direkter Schätzer für eine große übergeordnete Schicht, die mehrere kleine interessierende Areas oder Domains enthält, verfügbar ist und dazu verwendet wird, einen indirekten Schätzer für die Parameter der interessierenden Schicht zu gewinnen. Dabei muss die wesentliche Annahme getroffen werden, dass die kleine Gesamtheit die gleiche Charakteristik aufweist, wie die große, aus der der direkte Schätzwert generiert wurde. Das nationale Zentrum für Gesundheitsstatistik (NCHS) in den USA leistete im Jahr 1968 wesentliche Pionierarbeit für den Einsatz synthetischer Schätzer (vgl. Rao, 2003, S. 46). Direkte Schätzer konnten damals wegen teils zu geringen Stichprobenumfängen und den damit verbundenen hohen Standardfehlern nicht eingesetzt werden. Synthetische Schätzer können per Definition mit oder ohne Hilfsvariablen Anwendung finden. Die Tragweite der direkten bzw. indirekten Schätzmethode bleibt davon allerdings unberührt. Vorteile der synthetischen Schätzung sind sicherlich die einfache und intuitive Anwendbarkeit, die Option unterschiedliche Stichprobendesigns verwenden zu können sowie der elementare Gedanke der Small Area-Schätzung, von vergleichbaren Areas oder Domains Informationen zu erhalten. In diesem Zusammenhang wurde der bedeutsame und häufig verwendete Ausdruck *to borrow strength* geprägt (vgl. Rao, 2003, S. 2).

Ausgehend von der Eigenschaft der direkten Schätzer design-unverzerrt zu sein sowie von der Stärke der synthetischen Schätzer, in Form der in der Regel relativ zu direkten Schätzern geringeren Varianz, verspricht man sich von einer geeigneten Kombination dieser beiden eine gewinnbringende Symbiose. Wenngleich die synthetischen Schätzer in der Regel design-verzerrt sind, weisen sie meist geringere Variabilität auf. Mit Hilfe einer sinnvollen Verknüpfung erwartet man sozusagen einen Schätzer, der eine geringere Verzerrung als der synthetische Schätzer, gleichzeitig aber eine geringere Variabilität als der direkte Schätzer, aufweist. Die Frage nach der Gewichtung der beiden Schätzer in der zusammengesetzten Form kennt in erster Linie zwei Antworten. Einerseits wird die Variante einer stichprobengrößenabhängigen Kombination in der Literatur beschrieben, andererseits die Überlegung, einen Gewichtungsfaktor zwischen 0 und 1 zu suchen, bei dem der mean squared error (MSE) des zusammengesetzten Schätzers minimal wird (vgl. Rao, 2003, S. 57 ff.).

Während die bislang beschriebenen Verfahren der Parameterschätzung relativ leicht in Intuition und Umsetzung zu verstehen waren, handelt es sich bei den speziellen Small Area-Modellen um teilweise sehr komplexe Anwendungen der Small Area-Schätzung. Ein wesentlicher Unterschied zu den bereits beschriebenen Verfahren liegt in den expliziten Modellannahmen, die einen erklärenden Beitrag zu den Unterschieden und der Variabilität zwischen den Small Areas liefern sollen. Aufgrund dieser Modelle wird es deswegen möglich sein, Schätzwerte für Schichten zu erhalten, für die nur wenig bzw. sogar keine Stichprobendaten vorliegen. Aufgrund von Hilfsvariablen oder Zielvariablen aus anderen Small Areas ist es mit Hilfe des erklärten Zusammenhangs der Areas untereinander möglich, Schätzwerte zu generieren.

In der Literatur unterscheidet man zahlreiche Arten von Modellen, die zur Erklärung herangezogen werden können. Wenngleich einerseits in der Vielzahl und Flexibilität der Modellierung ein großer Vorteil der Small Area-Schätzung zu sehen ist, stellt diese Tatsache gleichzeitig die Schwierigkeit dar, diese Thematik richtig zum Einsatz zu bringen. Anhand einer entsprechenden Modelldiagnostik können Modelle evaluiert und gegebenenfalls angepasst und verbessert werden, um optimierte Schätzwerte auf den Small Areas zu erhalten. Ein weiterer Vorteil dieser Modelle ist, dass Area-spezifische Präzisionsmaße ermittelt werden können, die ebenso zur Beur-

teilung des Schätzverfahrens herangezogen werden können. Häufig vorkommende Modellvariationen liegen in der beliebigen problemadäquaten Ergänzung um fixe und zufällige Effekte, die den Erklärungsgehalt detaillierter spezifizieren können sowie in der Anzahl der betrachteten Ebenen. In diesem Kontext spricht man von so genannten Mehrebenenmodellen oder multi-level-Modellen (vgl. Longford, 1993, und Goldstein, 1995). Im Zusammenhang mit modellbasierten Schätzverfahren spricht man in der Literatur von Empirical Best Linear Unbiased Predictor, Empirical Bayes Method oder Hierarchical Bayes Method. Die Charakteristika dieser Methoden liegen vor allem in der Art und Weise des Auffindens der Lösung der linearen bzw. nichtlinearen Gleichungssysteme sowie in speziellen Modellanpassungen und Skalierungsproblemen.

Weitere Informationen und Details zu Small Area-Methoden sind in Rao (2003) wie auch Longford (2005) zu finden.

Einsatzmöglichkeiten der Small Area-Schätzung im Zensus in Deutschland

Das Potenzial der Small Area-Schätzung kommt schon dadurch zum Ausdruck, dass die Methodik von verschiedenen nationalen Statistikämtern in unterschiedlicher Intensität verfolgt und eingesetzt wird. Die Beispiele für Länder, die bereits solche Verfahren zum Einsatz bringen, sind zahlreich. Doch darüber hinaus beschäftigen sich immer mehr Institutionen mit dieser Thematik, um damit anstehende zukünftige Schätzprobleme zu lösen. Untersuchungen der beiden genannten Forschungsprojekte EURAREA und DACSEIS, begleitet von ausführlichen Simulationsrechnungen, haben gezeigt, dass Small Area-Verfahren sicherlich bei ausreichend Wissen und Erfahrung mit Gewinn eingesetzt werden können.

In DACSEIS (vgl. Münnich et al., 2004) wurden unterschiedliche Small Area-Verfahren, parallel zu Standard-Schätzmethoden (Horvitz-Thompson-Schätzer und Regressionsschätzer) evaluiert. Dabei verglich man zwischen unterschiedlichen Small Area-Kategorisierungen, unterschiedlichen Hilfsvariablen-Sets wie auch zwischen verschiedenen Untersuchungsvariablen. Zusätzlich variierte man die Simulationsrechnungen im Hinblick auf verschiedene Antwortausfallquoten. Bei den verwendeten Small Area-Verfahren unterschied man zwischen einem so genannten area-level-Modell und einem unit-level-Modell. Dabei handelt es sich um multilevel-Modelle (vgl. Longford, 1993, und Goldstein, 1995) mit erklärenden Effekten auf unterschiedlichen regionalen Niveaus. Darüber hinaus kombinierte man beide Modelle mittels einer Linearkombination mit mindestens approximativ design-unverzerrten Schätzern zu so genannten Empirical Best Linear Unbiased Predictors. Abhängig von Antwortausfallquoten, Qualität und Erklärungsgehalt der Hilfsvariablen sowie Art der Small Area-Kategorisierung konnte man teils sehr gute Ergebnisse erzielen – nicht nur im Hinblick auf Punkt- und Varianzschätzer sondern auch in Bezug auf die Wiedergabe gewisser systematischer Unterschiede zwischen den Small Areas (Disparitäten). Details zu Simulationsrechnungen und -variationen gehen aus Münnich et al., 2004, S. 23 ff. hervor.

Im Folgenden soll nun untersucht werden, inwiefern sich Small Area-Verfahren dazu eignen, spezielle Schätzprobleme im Rahmen eines registergestützten Zensus lösen zu können. Die Fragestellungen, Besonderheiten und Problempunkte sind sehr vielseitig und bedürfen einer nachhaltigen und genauen Analyse. Im ersten Teil dieses Beitrags konnte man bereits einen Eindruck dafür gewinnen, wo man besondere Sorgfalt walten lassen muss, um die Qualität der Ergebnisse nicht in Gefahr zu bringen. Es wird getestet werden müssen, welchen qualitativen

Standard man mit Small Area-Techniken erreichen kann. Bisherige Erfahrungen, auch in anderen Zusammenhängen, zeigen durchaus vorteilhafte Anwendungen. Doch es wird unabdingbar sein, speziell im Falle eines registergestützten Zensus die Anwendbarkeit zu überprüfen. Denkbar ist ein Einsatz sowohl bei der Schätzung der Fehlbestände und Karteileichen wie auch bei weiteren interessierenden Variablen. Die Small Area-Schätzung erscheint hier sicherlich aufgrund der Problemsituation, beispielsweise durch kleine Stichprobenumfänge oder Schätzungen für kleine Subpopulationen, als geeignet.

Neben der Untersuchung auf einen möglichen Einsatz spezieller Small Area-Verfahren bedarf es zusätzlicher intensiver Nachforschungen bezüglich des Stichprobendesigns. Es ist wichtig, die Schätzproblematik im Ganzen zu betrachten und somit das Stichprobendesign auf die zu verwendenden Schätzverfahren anzupassen. Des Weiteren sind die Verfahren darauf abzustimmen, welche Variablen bei welchem Informationsstand zu schätzen sind. Nachdem die Small Area-Verfahren in der Regel mit Hilfsinformationen arbeiten, muss geklärt sein, auf welche Daten zugegriffen werden kann (zum Beispiel Daten der Einwohnermeldeämter oder Daten der Bundesagentur für Arbeit). Die Qualität der Schätzergebnisse hängt in starkem Maße von der Güte der Hilfsinformationen ab. Grundsätzlich ist ein optimaler Einsatz von Small Area-Verfahren auf valide Angaben über die Strukturen der Informationen angewiesen. Das Stichprobendesign spielt hier eine nicht zu unterschätzende Rolle. Dabei kann das eigentliche Zensus-Ziel, die Ermittlung von Bevölkerungszahlen auf sachlich und räumlich untergliederten Niveaus und der damit verbundenen Schätzung von Karteileichen und Fehlbeständen sowie der Schätzung zusätzlicher Variablen, durchaus zu konfliktären Ergebnissen in Bezug auf die Qualität der Schätzungen führen.

Die Untersuchungen zu dieser Thematik sind noch in der Anfangsphase, weswegen noch keine verbindlichen Aussagen über eine zuverlässige Anwendung der Small Area-Verfahren getroffen werden können. Erste Testsimulationen zeigen aber durchaus viel versprechende Ergebnisse. Des Weiteren kann auf andere Länder, allen voran die USA, verwiesen werden, die bereits intensiv diese Methodik mit Erfolg zum Einsatz bringen.

Literatur

Fay, Robert E.; Herriot, Roger A. (1979): Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. In: *Journal of the American Statistical Association*, 74, S. 269 – 277.

Ghosh, Malay; Rao, Jon N.K. (1994): Small Area Estimation: An Appraisal. In: *Statistical Science*, 9, S. 55 – 93.

Goldstein, Harvey (1995): *Multilevel statistical models*. Second edition. London: Arnold. XIV + 178 S.

Lahiri, Partha (2005): Workshop 15. – 16. März 2005: Introduction to Small Area Estimation. ZUMA Mannheim 2005. Unveröffentlichtes Manuskript.

Longford, Nicholas T. (1993): Random Coefficient Models. Oxford: Oxford University Press. (Oxford statistical science series; 11) XIV + 270 S.

Longford, Nicholas T. (2005): Missing data and small-area estimation: modern analytical equipment for the survey statistician. Heidelberg: Springer. (Statistics for social science and public policy) XV + 357 S.

Münnich, Ralf; Magg, Kersten; Söstra, Kaja; Schmidt, Katrin; Wiegert, Rolf (2004): Workpackage 10: Variance Estimation for Small Area Estimates: Deliverables 10.1 and 10.2. URL <http://www.dacseis.de>. - IST-2000-26057-DACSEIS Reports.

Münnich, Ralf; Schmidt, Katrin (2002): Small Area Estimation in der Bevölkerungsstatistik. In: Baden-Württemberg in Wort und Zahl, 3/2002, S. 139 – 145.

Nirel, Ronit; Glickman, Hagit; Ben-Hur, Dan (2003): A Strategy for a System of Coverage Samples for an Integrated Census. In: Proceedings of Statistics Canada Symposium 2003: Challenges in Survey Taking for the Next Decade.

Pfeffermann, Danny (2005): Small Area Estimation Using Time Series Models Subject to Benchmarking Constraints. In: Proceedings of SAE 2005 Conference: Challenges in Statistics Production for Domains and Small Areas: Scientific Conference Featuring Main Results of, and Issues Raised by, the EURAREA Project. Jyväskylä, Finland.

Rao, Jon N.K. (2003): Small Area Estimation. Hoboken: John Wiley & Sons. (Wiley series in survey methodology) XXIII + 313 S.

Renaud, Anne (2004): Swiss Federal Statistical Office (Hrsg.): Methodology Report: Coverage Estimation for the Swiss Population Census 2000: Estimation Methodology and Results. Neuchâtel: Swiss Federal Statistical Office. (Series: Swiss Statistics) 147 S.

Schaible, Wesley L. (1996) (Hrsg): Indirect Estimators in U.S. Federal Programs. Heidelberg: Springer. (Lecture Notes in Statistics; Vol. 108) 212 S.

Autoren- und Kontaktangaben:

Dipl.-Volksw. Dipl.-Betriebsw. (BA) Kersten Helge T. Magg, Eberhard-Karls Universität Tübingen, Wirtschaftswissenschaftliche Fakultät, Abteilung Statistik, Ökonometrie und Unternehmensforschung, E-Mail: kersten.magg@uni-tuebingen.de

Prof. Dr. Ralf Münnich, Universität Trier, Fachbereich IV, Lehrstuhl für Wirtschafts- und Sozialstatistik, E-Mail: muennich@uni-trier.de

Josef Schäfer, Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen, Referat 332, Interfachliche Erhebungen, Mathematisch-wissenschaftliche Aufgaben, E-Mail: josef.schaefer@lds.nrw.de

Zum Einfluss von Marktfaktoren, institutionellen und individuellen Faktoren auf Jugendarbeitslosigkeit in Westeuropa*)

Analysen mit Zeitreihen und Paneldaten des europäischen Haushaltspanels

Katrin Golsch

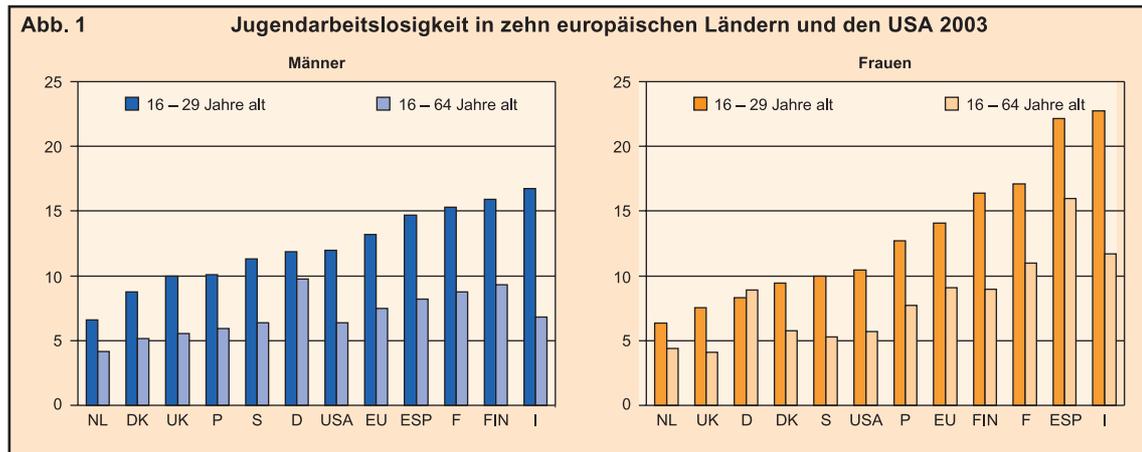
In den meisten europäischen Ländern haben sozioökonomische Veränderungen im Arbeitsmarkt und Wohlfahrtsstaat Marktrisiken für individuelle Akteure erhöht sowie den Grad sozialer Sicherung verringert, und zwar insbesondere für junge Arbeitnehmerinnen und Arbeitnehmer, die über wenig Arbeitsmarkterfahrung und geringe Bindungen an den Arbeitsmarkt verfügen. Ziel dieses Beitrags ist es, den Einfluss von Marktfaktoren, institutionellen und individuellen Faktoren auf das Arbeitslosigkeitsrisiko von jungen Erwachsenen in zehn westeuropäischen Ländern zwischen 1994 und 2001 zu untersuchen. Die Ergebnisse einer Analyse von Zeitreihen und Paneldaten des europäischen Haushaltspanels eröffnen neue Einblicke in die internationale Variation der Jugendarbeitslosigkeit. Die beobachteten länderspezifischen Unterschiede sind auf Marktfaktoren und institutionelle Unterschiede in den untersuchten Ländern zurückzuführen. Soziale Ungleichheit lässt sich überproportional häufig in bestimmten Gruppen von jungen Erwachsenen wie zum Beispiel Geringqualifizierten und unerfahrenen Arbeitern mit fragmentierten Karrieren feststellen.

1 Einleitung

Eine Vielzahl von Forschungsbeiträgen hat sich in der Vergangenheit mit der Analyse der Integration von jungen Erwachsenen in den Arbeitsmarkt befasst. Diese Studien diskutieren sowohl die Mechanismen sozialer Ungleichheit als auch den Einfluss von verschiedenen länderspezifischen Arrangements in Bildung, Arbeitsmarkt und Wohlfahrtsstaat (Allmendinger, 1989; Breen, 2005; Gangl, 2001, 2003; Shavit und Müller, 1998; Russell and O'Connell, 2001). In den letzten Jahren werden eine wachsende Interdependenz der regionalen und nationalen Ökonomien, eine Intensivierung der Konkurrenz, industrielle Restrukturierung und Technologiewandel sowie eine wachsende Instabilität der Arbeits-, Kapital- und Produktmärkte beobachtet. Diese globalen Veränderungen, so wird häufig argumentiert, führen zu einer wachsenden ökonomischen, sozialen und vor allem temporalen Unsicherheit im Arbeitsleben (Blossfeld u. a., 2005; Golsch, 2005). Bettet man diese Entwicklungen in einen größeren Kontext der zunehmenden Deregulierung und Flexibilisierung der Arbeitsmärkte sowie einschneidenden Kürzungen in verschiedenen Gebieten der Wohlfahrtsstaatsversorgung während der 1980er und 1990er Jahre, so kann wachsende Arbeitsmarktunsicherheit als ein besonderes Dilemma für junge Erwachsene, die nach Arbeit suchen, und junge Arbeitnehmerinnen und Arbeitnehmer mit mangelnder Erfahrung und nur geringer Bindung zum Arbeitsmarkt gesehen werden (Golsch, 2005).

*) Die vorliegende Untersuchung wurde während eines durch das EuroPanel Users' Network (EPUNet) finanzierten Forschungsaufenthaltes am Deutschen Institut für Wirtschaftsforschung durchgeführt. Die Finanzierung des EPUNet erfolgt aus Mitteln der EU innerhalb des Programms 'Improving the Human Research Potential and the Socio-Economic Knowledge Base' (Vertragsnummer HPSE-CT-2002-50021).

Abbildung 1 zeigt das Ausmaß der Jugendarbeitslosigkeit in zehn europäischen Ländern und den USA und liefert den Hintergrund vor dem diese Studie motiviert ist. Egal ob Jugendarbeitslosigkeit oder andere Formen unsicherer Arbeitsmarktpositionen in den Blick genommen werden, das Bild ist eindeutig. Insbesondere junge Arbeitnehmerinnen und Arbeitnehmer zwischen 16 und 29 Jahren sind mit den größten Risiken wachsender Arbeitsmarktunsicherheit konfrontiert. Die Risiken prekärer Beschäftigung und sozialer Exklusion variieren jedoch zwischen den einzelnen europäischen Ländern beträchtlich. Aus Abbildung 1 können noch keine Rückschlüsse über die Ursachen dieser länderspezifischen Unterschiede gezogen werden, da die Marktfaktoren und institutionellen Faktoren in vielfacher Weise variieren.



Bislang liegen nur wenige systematische und international vergleichende Studien vor, die den Einfluss von Marktfaktoren und institutionellen Faktoren auf die Integration in den Arbeitsmarkt analysieren (Breen, 2005). Es ist deshalb noch nicht hinreichend erforscht, inwiefern Marktfaktoren und institutionelle Faktoren auf der einen Seite und individuelle Erwerbsgeschichte, Bildung und Beschäftigungsverhältnis auf der anderen Seite Übergänge von jungen Erwachsenen in Arbeitslosigkeit beeinflussen.

Der vorliegende Beitrag diskutiert die Veränderungen von Marktfaktoren und institutionellen Faktoren im Zeitalter der Globalisierung und analysiert das Arbeitslosigkeitsrisiko von Männern und Frauen im Alter von 16 bis 29 Jahren zwischen 1994 und 2001. Die vergleichende Analyse stützt sich auf zehn Länder aus insgesamt vier Wohlfahrtsstaatsregimen und bildet somit einen Großteil der institutionellen Heterogenität von westeuropäischen Ländern ab. Deutschland, die Niederlande und Frankreich repräsentieren das konservative Regime; Dänemark, Finnland und Schweden stehen für die sozialdemokratischen Regimes. Die familienorientierten Regimes werden durch Spanien, Italien und Portugal vertreten; das Vereinigte Königreich spiegelt den liberalen Regimetyp wider. Die Untersuchung verwendet Zeitreihen verschiedener Quellen sowie Paneldaten des Europäischen Haushaltspanels (ECHP) und sucht Antworten auf die folgenden drei Forschungsfragen:

- In welchem Ausmaß sind junge Arbeitnehmerinnen und Arbeitnehmer in diesen Ländern einem größeren Arbeitslosigkeitsrisiko ausgesetzt?
- Welche Rolle spielen Markt- und institutionelle Faktoren?
- Inwiefern erwächst soziale Ungleichheit aufgrund von interindividuellen Unterschieden hinsichtlich Bildungsniveau, Erwerbsgeschichte oder Beschäftigungsverhältnis?

Der nächste Abschnitt entwickelt den theoretischen Rahmen der Studie. Es folgt eine Beschreibung der Datenquellen, statistischen Methoden und Variablen. Den Hauptteil stellt die empirische Analyse von Jugendarbeitslosigkeit dar. Der Beitrag schließt mit einer Zusammenfassung der zentralen Ergebnisse und einigen Schlussfolgerungen.

2 Theoretische Vorüberlegungen

In den letzten Jahren hat sich die Arbeitswelt aufgrund der Intensivierung des internationalen Wettbewerbs, der stärkeren Instabilität der Arbeits-, Kapital- und Produktmärkte sowie der tiefgehenden Transformationen der Produktionsprozesse und des immer rascheren beruflichen und sektoralen Strukturwandels gewandelt (Mills and Blossfeld, 2005; Heery and Salmon, 2000; Ladipo and Wilkinson, 2002). Dieser Wandel manifestiert sich vor allem in der wachsenden Nachfrage nach flexiblen Arbeitskräften, der Forderung nach einer stärkeren Flexibilisierung der Arbeitsbeziehungen und den Veränderungen der den Berufseinsteigern offerierten Erwerbsformen und Beschäftigungsverhältnissen (Zunahme von befristeten Beschäftigungen und Teilzeitarbeit). Gleichzeitig dehnt sich in den heutigen Wissensgesellschaften die Ausbildungsphase immer weiter aus und die einzelnen Akteure sind zunehmend gezwungen, sich kontinuierlich weiterzubilden (Blossfeld, 2002). Ein Trend zu verschärfter Konkurrenz zwischen den Nationalstaaten und Wohlfahrtsstaatsabbau sind ebenfalls zu beobachten (Esping-Andersen, 1996). In vielen europäischen Ländern haben diese sozioökonomischen Veränderungen die Marktrisiken verschärft und die soziale Sicherung insbesondere für junge Erwachsene, die gerade in den Arbeitsmarkt eintreten, verringert (Golsch, 2005).

Die verschiedenen europäischen Länder haben jedoch aufgrund ihrer länderspezifischen institutionellen Kontexte und historischen Pfadabhängigkeiten sehr unterschiedlich auf diese globalen Veränderungen reagiert (Blossfeld, 2002; Esping-Andersen, 1999; Mayer, 1997). So bestehen beispielsweise länderspezifische Variationen hinsichtlich des Ausmaßes zu dem es nationalen Bildungssystemen gelingt, sich an die Herausforderungen expandierender globaler Märkte anzupassen (Regini, 1997). Gleichzeitig unterscheiden sich die Beschäftigungssysteme. Während einige Länder die Arbeitsmärkte flächendeckend flexibilisiert und dereguliert haben, tritt Arbeitsmarktflexibilisierung in anderen Ländern nur als ein kontrolliertes Experiment auf, so dass Unternehmen häufig nur unzureichend auf zunehmende Flexibilisierungszwänge reagieren können (Regini, 2000). Ein weiterer Faktor ist das Ausmaß zu dem einige Wohlfahrtsstaaten die sozialen Risiken mildern wohingegen in anderen Ländern aufgrund rudimentärer Wohlfahrtsstaatsversorgung diese Risiken für einzelne Akteure sogar verschärft werden (Esping-Andersen, 1993). Individuelle Lebensverläufe sind in diesen länderspezifischen institutionellen Kontexten eingebettet, so dass individuelle Antworten auf zunehmende Marktrisiken und abnehmende soziale Sicherung zwischen den Ländern variieren.

In der Analyse von Jugendarbeitslosigkeit in Westeuropa spielen jedoch nicht nur Marktfaktoren und institutionelle Faktoren eine große Rolle. Auf der individuellen Ebene sind auch signifikante Unterschiede zwischen individuellen Akteuren, die in dem gleichen institutionellen Kontext leben, zu erwarten. Ein größeres Arbeitsloskeitsrisiko wird insbesondere für Arbeitnehmerinnen und Arbeitnehmer mit geringer Bildung, mangelnder Arbeitsmarkterfahrung und unsicherer Jobposition erwartet. Die Diskussion wendet sich nun den spezifischen Forschungshypothesen dieses Beitrags zu.

Hypothesen über die Rolle von Marktfaktoren und institutionellen Faktoren

Die vorliegende Studie betrachtet die folgenden vier länderspezifischen Marktfaktoren und institutionellen Faktoren:

1) Ökonomische Entwicklung und Handelsöffnung

Handelsöffnung, und somit Öffnung für globale Konkurrenz, führt zu niedrigeren Arbeitslosigkeitsraten. Während in Zeiten des ökonomischen Aufschwungs das Arbeitslosigkeitsrisiko geringer ist.

2) Bildungssystem und Bildungsanforderungen

In Ländern mit einem hohen Anteil an hochqualifizierten Erwerbspersonen, wie beispielsweise in den skandinavischen Ländern aber auch in Deutschland und Frankreich, werden geringere Jugendarbeitslosigkeitsraten erwartet. In diesem Zusammenhang ist es von Bedeutung ob, und in welchem Ausmaß, es nationalen Bildungssystemen gelingt, die Herausforderungen eines expandierenden globalen Marktes zu erfüllen (Regini, 1997). Mit der wachsenden Nachfrage nach vielseitig qualifizierten und erfahrenen Arbeitskräften konkurrieren Jobanfänger/-innen nicht nur in wachsendem Maße mit denjenigen, die bereits beschäftigt sind – was ihre Aussichten auf eine Beschäftigung zunehmend verschlechtert. In den heutigen Wissensgesellschaften sind diese jungen Erwachsenen auch zunehmend gezwungen, ihre Fähigkeiten kontinuierlich zu erweitern (Blossfeld, 2002). Ein weiterer wichtiger Aspekt, der jedoch nicht im Rahmen dieses Beitrags behandelt werden kann, sind die länderspezifischen Unterschiede des Bildungs- und Ausbildungssystems, die in einigen Ländern komparative Vorteile, in anderen Ländern aber auch Nachteile bezüglich der Wartezeit auf und der Stabilität einer ersten Beschäftigung bergen (Blossfeld and Stockmann, 1998/1999). Bisherige Forschungen haben gezeigt, dass Bildungssysteme sowohl einen signifikanten Einfluss auf den Zeitpunkt des Arbeitsmarkteintritts als auch den Rekrutierungsprozess haben. Während die Zunahme von Beschäftigungsinstabilität zu Beginn der Erwerbskarriere besonders auffällig in den süd-europäischen Ländern ist, nehmen die nordeuropäischen Länder im europäischen Vergleich eher eine mittlere Position ein und der Übergang in den primären Arbeitsmarkt scheint in den zentraleuropäischen Ländern deutlich weniger durch ökonomische, soziale und vor allem temporale Unsicherheiten gekennzeichnet zu sein (Blossfeld u. a., 2005).

3) Offenheit des Arbeitsmarktes

Auf der einen Seite bedeuten geringe gewerkschaftliche Bindung, wenig verbreitete Flächentarifverträge und ein schwacher Beschäftigungsschutz in offenen Beschäftigungssystemen wie z. B. im Vereinigten Königreich weniger Sicherheit auf der Arbeitnehmerseite. In einem wachsenden globalen Markt sind Arbeitnehmerinnen und Arbeitnehmer weniger geschützt gegenüber Marktrisiken als Arbeitnehmerinnen und Arbeitnehmer in Ländern mit eher geschlossenen Beschäftigungssystemen. Auf der anderen Seite ist es ein wichtiges Charakteristikum der offenen Systeme, dass eine schwache institutionelle Regulierung weniger Risiken für Arbeitgeber birgt und die Kosten von Neueinstellungen und Kündigungen senkt. Dies bedeutet auch, dass es im Vergleich zu den stärker regulierten Ländern wie beispielsweise Deutschland und Spanien weniger Hürden für diejenigen gibt, die versuchen, eine erste Beschäftigung zu finden. Es wird deshalb erwartet, dass das Risiko, arbeitslos zu werden, in geschlossenen Arbeitsmärkten größer ist als in offenen.

4) Härte des Beschäftigungsschutzes

Ein restriktives und rigides Beschäftigungssystem sichert individuelle Akteure hinsichtlich der Risiken des Job- oder Einkommensverlustes ab. Eine Flexibilisierung des Beschäftigungsschutzes stellt dieses balancierte System, das auf langfristigen, kooperativen Beziehungen zwischen Arbeitgebern und Arbeitnehmern fußt, in Frage. Während eine eher restriktive Beschäftigungsschutzpolitik Unternehmen wenige Möglichkeiten lässt, auf stetig wachsende Flexibilisierungszwänge zu reagieren, eröffnen deregulierende Maßnahmen neue Optionen. In dem Maße in dem Arbeitsmarktflexibilität Kündigungen, Gehaltskürzungen oder befristete Beschäftigung erleichtert werden diese Maßnahmen auch den Grad der Sicherheit, auf den sich Arbeitnehmerinnen und Arbeitnehmer verlassen können, verringern.

Auf der anderen Seite kann ein rigides System auch eine signifikante Barriere für diejenigen darstellen, die versuchen in den Arbeitsmarkt (wieder-)einzutreten – vor allem für junge Erwachsene, die nach Arbeit suchen. Wenn es diesen Personen nicht gelingt, eine unbefristete Position zu erlangen, die mit ihren individuellen Fähigkeiten übereinstimmt, dann ist es wahrscheinlich, dass sie längerfristig „Outsider“ bleiben. Im Unterschied dazu werden diejenigen, die bereits auf unbefristeten Positionen beschäftigt sind, gute Karriere- und Weiterbildungsperspektiven eröffnet. Langfristige Beschäftigungsverhältnisse für diese „Insider“ gehen daher oftmals mit Langzeitarbeitslosigkeit und Inaktivität von „Outsidern“ einher. Arbeitsmarktungleichheit und soziale Exklusion wird so verschärft. Zusammenfassend wird in diesem Beitrag davon ausgegangen, dass im Vergleich zu den eher flexiblen Systemen in dem Vereinigten Königreich und Dänemark Übergänge in Arbeitslosigkeit wahrscheinlicher sind für unerfahrene Arbeitnehmerinnen und Arbeitnehmer in Ländern mit einem strikten Beschäftigungsschutz wie beispielsweise den südeuropäischen Ländern.

Hypothesen über die Rolle von individuellen Faktoren

Der gegenwärtige Wandel der Arbeitsmärkte und institutionellen Kontexte hat in Abhängigkeit von den persönlichen Ressourcen unterschiedliche Implikationen für individuelle Erwerbskarrieren. Die spezifischen Individualhypothesen dieses Beitrags sind:

1) Bildung

Mit dem Übergang zu mehr qualifizierten Jobs und der zunehmenden Bedeutung von Bildungszertifikaten werden insbesondere diejenigen von Unsicherheit betroffen sein, die nur unzureichend für den Übergang vom Bildungssystem in den Arbeitsmarkt gerüstet sind. Dieser Beitrag nimmt deshalb an, dass Arbeitnehmerinnen und Arbeitnehmer mit hohen Bildungsabschlüssen seltener von Arbeitslosigkeit betroffen sein werden als Geringqualifizierte.

2) Beschäftigungsverhältnis

Aufgrund der zunehmenden Instabilität von Arbeits-, Kapital-, und Produktmärkten und der Intensivierung der globalen Konkurrenz sind Firmen gezwungen, sich schnell und effizient an sich stetig wandelnde Marktgegebenheiten anzupassen. Als eine Konsequenz daraus zögern Arbeitgeber zunehmend, neue Arbeitnehmerinnen und Arbeitnehmer auf unbefristeten Positionen einzustellen und neigen eher dazu, Neueinsteigern befristete Stellen zu offerieren, um deren Arbeitspotenzial zu erproben (Rekrutierungsstrategie) und die zu erwartenden Kosten im Falle einer Fehleinstellung zu reduzieren (Hire-and-Fire-Strategie). Es wird deshalb angenommen, dass junge Arbeitnehmer in befristeten Jobs oder marginaler Teilzeitarbeit häufiger von Arbeitslosigkeit bedroht sind.

3) Erwerbsgeschichte

Die Integration in den Arbeitsmarkt ist ein schrittweiser Prozess, der durch Phasen der Arbeitslosigkeit und Nichtbeschäftigung, Beförderung sowie berufliche, innerbetriebliche und industrielle Mobilität charakterisiert ist. Mit fortschreitender Karriere wächst das Engagement, die Loyalität, die Arbeitsmarkterfahrung und grundlegendes berufs- und firmenspezifisches Wissen von Arbeitnehmerinnen und Arbeitnehmern. Dies ist von zentraler Bedeutung da Entlassungen im Vergleich zu weniger erfahrenen und trainierten Arbeitskräften für Arbeitgeber zunehmend kostspielig werden. Es wird daher erwartet, dass ein Zuwachs an Erfahrung das Arbeitslosigkeitsrisiko einer Person reduziert. Ein zweiter zentraler Aspekt der Erwerbsgeschichte ist Arbeitslosigkeitserfahrung. Hier wird argumentiert, dass fragmentierte Karrieren ein schlechtes Signal für Arbeitgeber darstellen und somit das Risiko einer erneuten Arbeitslosigkeit erhöhen.

Synthese

Schließlich wird davon ausgegangen, dass Marktfaktoren und institutionelle Faktoren auf der einen Seite sowie Bildung und vorangegangene Erwerbsgeschichte auf der anderen Seite miteinander interagieren. Dieser Beitrag befasst sich mit drei Forschungshypothesen über die spezifischen Dilemmata auf der individuellen Ebene:

1) Bildungsnachfrage und individuelle Fähigkeiten

Die beschriebenen Veränderungen des Arbeitsmarktes verstärken die Nachfrage nach flexiblen, hochqualifizierten Arbeitskräften. Als Konsequenz daraus wird es für junge Erwachsene, die die nötige Qualifikation nicht vorweisen können, zunehmend schwieriger mit anderen Arbeitskräften zu konkurrieren. Die Bildungsnachfrage variiert jedoch nach wie vor zwischen den westeuropäischen Ländern. Es wird deshalb erwartet, dass in Ländern, in denen die Bildungsnachfrage groß ist, die Marktrisiken am größten für Geringqualifizierte sind.

2) Offenheit des Arbeitsmarktes und individuelle Fähigkeiten

Ein wichtiger Aspekt in der Analyse der Arbeitsmarktintegration sind länderspezifische Unterschiede der Beschäftigungssysteme. Unsicherheit ist eher gleich verteilt unter den Arbeitnehmerinnen und Arbeitnehmern, die in einem offenen System leben, aber die Möglichkeiten in den Arbeitsmarkt (wieder-)einzutreten sind gleichzeitig größer als in stärker regulierten Systemen. Im Unterschied dazu sind Arbeitnehmerinnen und Arbeitnehmer in einem geschlossenen System weitestgehend geschützt gegenüber neuen Marktrisiken, gleichzeitig bildet dieses System aber Barrieren für all diejenigen, die versuchen in den Arbeitsmarkt einzutreten. Die Konsequenz ist, dass in beiden Systemen junge Erwachsene mehr Unsicherheiten ausgesetzt sind als erfahrene Arbeitskräfte. Es wird deshalb davon ausgegangen, dass der Unterschied zwischen hochqualifizierten und geringqualifizierten Arbeitnehmerinnen und Arbeitnehmern in Ländern mit einem offenen oder geschlossenen Beschäftigungssystem wenig ausgeprägt ist.

3) Beschäftigungsschutz und Erwerbskarriere

In Ländern mit deregulierten Arbeitsmärkten werden Marktrisiken von Arbeitgebern an Arbeitnehmerinnen und Arbeitnehmern weitergeleitet. Im Unterschied dazu, wird die betriebliche Anpassung an sich stetig wandelnde Marktgegebenheiten durch einen rigiden Arbeitsschutz erschwert. Unternehmen fehlt die Möglichkeit, je nach Marktlage Arbeitnehmerinnen und Ar-

beitnehmer einzustellen und zu entlassen. Dies setzt Unternehmen unter Druck, das Arbeitspotenzial von neuen Arbeitnehmerinnen und Arbeitnehmern sorgfältig zu überprüfen, und in dieser Hinsicht werden vorangegangene Phasen der Arbeitslosigkeit dem Arbeitgeber ein schlechtes Signal geben. Im Vergleich zu Ländern mit schwachem Beschäftigungsschutz sind deshalb Arbeitnehmerinnen und Arbeitnehmer mit fragmentierten Karrieren in Ländern mit einem starken Beschäftigungsschutz häufiger von Arbeitslosigkeit betroffen.

3 Daten, Methoden und Variablen

Datenquellen und Stichprobe

Die vorliegende Analyse basiert auf einer Verknüpfung von Zeitreihen und Haushaltspaneldaten für die zehn ausgewählten Länder. Dabei wurde auf unterschiedliche Quellen wie beispielsweise die OECD Arbeitsmarktstatistik Datenbasis und ADB Datenbasis sowie Eurostats New Cronos Datenbasis zurückgegriffen und der Zeitraum von 1994 bis 2001 abgebildet. Die Individualdaten wurden aus dem ECHP extrahiert (1994 – 2001).¹⁾ Das ECHP wurde als eine harmonisierte Längsschnittuntersuchung konzipiert, um Informationen über Beschäftigung, Einkommen, bisherige Jobs, Arbeitsplatzsuche, Bildung und Zufriedenheit mit verschiedenen Aspekten des Lebens und der Lebenssituation in der EU zu erheben. 1994 wurde eine Stichprobe von ungefähr 60 500 repräsentativen Haushalten mit 130 000 Erwachsenen im Alter von 16 und mehr Jahren in den damals zwölf Mitgliedsstaaten befragt. Die gleichen Haushalte und ihre Mitglieder wurden in diesen Ländern bis 2001 weiterverfolgt. Seit 1995 bzw. 1996 wurden auch in Österreich und Finnland Interviews durchgeführt. 1997 wurde die Befragung der ECHP Stichprobe in Deutschland und England allerdings beendet. Stattdessen wurden das deutsche Sozioökonomische Haushaltspanel (GSOEP) und das britische Haushaltspanel (BHPS) herangezogen, um vergleichbare Daten für das ECHP zu generieren. Das Weiterverfolgungskonzept des ECHP sah vor, dass alle Mitglieder der Startstichprobe sowie alle neuen Haushalte weiterverfolgt werden. Dennoch variieren die Responseraten zwischen den teilnehmenden Ländern nicht unerheblich.²⁾

1. Beschreibung der Untersuchungspopulation					
Land	Individuen (Beobachtungen)	Panelwellen			
		Mittelwert	Std. Abw.	Min	Max
Dänemark	460 (1 614)	3,509	2,002	1	8
Niederlande	1 004 (3 074)	3,062	1,872	1	8
Frankreich	1 312 (4 467)	3,405	1,991	1	8
Italien	1 631 (5 713)	3,503	1,990	1	8
Spanien	1 829 (6 105)	3,338	1,932	1	8
Portugal	1 240 (4 236)	3,416	1,826	1	8
Finnland	973 (2 832)	2,911	1,557	1	6
Deutschland	639 (2 721)	4,258	1,911	1	8
UK	949 (3 689)	3,887	1,993	1	8
Gesamt	10 037 (34 451)				

Quelle: ECHP (1994 – 2001), eigene Berechnungen.

1) Weitere Details zum Fragebogen und den Erhebungsmethoden finden sich in European Commission Eurostat (1996) und auf der folgenden Website:

<http://forum.europa.eu.int/irc/dsis/echpanel/info/data/information.html> [17. 06. 05].

2) Mehr Informationen finden sich in Doc.Pan 92/97 and Doc.Pan 108/99:

(<http://forum.europa.eu.int/irc/dsis/echpanel/info/data/information.html> [17. 06. 05])

Ein besonderes Augenmerk wurde auf die Datenvergleichbarkeit gerichtet. Die folgenden Analysen konzentrieren sich auf eine unbalancierte Längsschnittstichprobe von Männern und Frauen im Alter von 16 bis unter 30 Jahren zwischen 1994 und 2001. Des Weiteren werden nur Mitglieder der Ausgangsstichprobe mit einem vollständigen Interview und ohne körperliche oder geistige Behinderungen, die zwischen 1994 und 2001 erstmalig in den Arbeitsmarkt eintreten, berücksichtigt.³⁾ Tabelle 1 zeigt einige deskriptive Statistiken für diese Untersuchungsgruppe.

Länderauswahl und Zeitrahmen

Wenngleich die Länderauswahl theoretisch fundiert ist, so kann die Forschungsstrategie von einem methodologischen Gesichtspunkt kritisch hinterfragt werden (Ebbinghaus, 2005). Erstens verwendet dieser Beitrag eine bewusste Auswahl von zehn europäischen Ländern und diese willkürliche Auswahl erfüllt nicht alle Kriterien der Inferenzstatistik. Zweitens ist die empirische Reichweite der Studie temporär und regional begrenzt. Die institutionelle Vielfalt kann nicht vollständig abgebildet werden und während die potenziellen Quellen von ökonomischer, sozialer und temporaler Unsicherheit sich kontinuierlich wandeln, bilden alte, historisch gewachsene institutionelle Arrangements gewisse Rigiditäten. Genau vor diesem Hintergrund will dieser Beitrag weitere Untersuchungen mit anderen Ländern und längeren Untersuchungszeiträumen motivieren.

Methoden

Die abhängige Variable in der Zeitreihenanalyse ist die Jugendarbeitslosigkeitsrate (16 bis unter 30 Jahre) nach Geschlecht in einem gegebenen Land und einem bestimmten Jahr. Aufgrund von Pfadabhängigkeiten auf der Aggregatebene sind sowohl die abhängige Variable wie auch die erklärenden Variablen häufig nicht-stationäre Prozesse und hohe serielle Autokorrelation ist zu erwarten (Kittel, 2005).

Die Zeitreihenanalyse schätzt Regressionsmodelle mit Zufallseffekten und fixen Effekten (Wooldridge, 2003). Für die Regressionsdiagnostik wird der Breusch und Pagan Lagrangian Multiplier Test und der Hausman Test verwendet. Anhand des ersten Tests wird überprüft, ob die Nullhypothese zutrifft, dass das unit-spezifische Residuum gleich Null ist und somit eine zentrale Annahme der Regressionsanalyse mit Zufallseffekten erfüllt ist. Der Hausman Test wird verwendet, um die fixen Effekte mit den Zufallseffekten zu vergleichen und die Nullhypothese zu überprüfen, dass die mit dem effizienten Zufallseffekteschätzer geschätzten Koeffizienten vergleichbar sind mit dem konsistenten Fix-Effekte Schätzer.

Für die Analyse der Individualdaten wird auf Verfahren der diskreten Ereignisanalyse zurückgegriffen (Allison, 1982). Die abhängige Variable in den Logitmodellen ist der Übergang in Arbeitslosigkeit zwischen zwei Panelwellen, kodiert mit 1 für einen Übergang und 0 für keinen Übergang. Alle Beobachtungen, in denen konkurrierende Risiken auftreten, werden von der Analyse ausgeschlossen.⁴⁾ Um unbeobachtete Heterogenität zu kontrollieren werden Logitmodelle mit Zufallseffekten geschätzt.

3) Schweden musste von den weiteren Analysen ausgeschlossen werden, da dort im Rahmen des ECHP nur Querschnittsdaten erhoben wurden. – 4) Diese konditionale Maximum-Likelihood Schätzung ergibt unverzerrte, konsistente und asymptotisch normalverteilte Punktschätzer (vg. Allison, 1995; Begg und Gray, 1984).

Variablen

Im Folgenden wird der Logarithmus der Arbeitslosenquote betrachtet, da die Verteilung schief ist und so die Daten näherungsweise normal verteilt sind. Die Variable \log (Arbeitslosigkeit) wird dann als eine Funktion von Marktfaktoren und institutionellen Faktoren sowie dem individuellen Bildungsstand, dem Beschäftigungsverhältnis und der Erwerbsgeschichte betrachtet. Die Variablen, die die Institutionen beschreiben, sind in Tabelle 2 zusammengefasst. Um zu überprüfen, ob die Varianz der institutionellen Variablen sowohl zwischen den Ländern als auch innerhalb der Länder über die Zeit hinreichend groß ist, zeigt Tabelle 2 auch einige deskriptive Statistiken für diese institutionellen Variablen und zerlegt die Standardabweichung in ihre einzelnen Komponenten. Für einige Länder stehen nicht für alle Zeitpunkte und alle Marktfaktoren bzw. institutionellen Faktoren Informationen zur Verfügung. Um diese unvollständigen Zeitreihen zu vervollständigen wurde die entsprechende Information für ein Land von dem vorangehenden Zeitpunkt kopiert. Da die Anzahl dieser Lücken nur klein ist und sich die institutionellen Faktoren nur sehr langsam wandeln, kann davon ausgegangen werden, dass dieses Vorgehen eine gute Approximation an die tatsächliche Entwicklung darstellt.

2. Operationalisierung der Marktfaktoren und institutionellen Faktoren und deskriptive Statistiken						
Operationalisierung und Quelle		N	Mittelwert	Std. Abw.	Min	Max
Wirtschaftliche Entwicklung	Bruttoinlandsprodukt pro Kopf, PPS, (EU25 = 0), 1995 – 2001; Stapel, S., Pasanen, J. und Reinecke, S. (2004) Prices and purchasing power parities, Statistics in focus. Economy and finance, 53/2004	140	9,54	gesamt: 14,73 zwischen: 14,93 innerhalb: 1,96	-27,00	27,00
Handelsoffenheit	Trade-to-GDP ratio (a), 1990, 1995, 1998, 1999; OECD, ADB database, May 2001 und IMF, Balance of Payments Statistics, zeitverzögert und zentriert	140	0,00	gesamt: 9,98 zwischen: 10,06 innerhalb: 1,61	-10,66	27,74
Bildungsanforderung	% der Bevölkerung (20 – 24) m. mindestens Sekundarstufe-II-Abschluss, 1994 – 2001; Eurostat (2004) New Cronos database, zeitverzögert und zentriert	140	0,00	gesamt: 12,83 zwischen: 12,54 innerhalb: 3,78	-32,00	18,01
Geschlossenheit des Arbeitsmarktes	Langzeitarbeitslose (12 Monate und länger) in % der Erwerbsbevölkerung nach Geschlecht, 1994 – 2001; Eurostat (2004) New Cronos database, zeitverzögert und zentriert	134	0,00	gesamt: 3,13 zwischen: 2,94 innerhalb: 1,09	-3,47	13,73
Härte des Beschäftigungsschutzes	EPL Wert (b), Version 1, 1990, 1998, 2003; OECD (2004) Labor Market Statistics – Indicators, zeitverzögert und zentriert	140	0,00	gesamt: 0,95 zwischen: 0,87 innerhalb: 0,42	-2,01	1,59

Anmerkungen:

(a) Der Trade-to-GDP Ratio ist definiert als der durchschnittliche Anteil der Exporte und Importe von Waren und Dienstleistungen am BIP. Dieser Indikator kann auch durch einige andere Faktoren beeinflusst werden wie z. B. die Ländergröße, Anzahl der Nachbarstaaten, Außenhandelspolitik, Ökonomiestruktur, oder die Präsenz von multinationalen Firmen. Da sich die vorliegende Studie jedoch auf westeuropäische Länder konzentriert, wird angenommen, dass diese Störfaktoren nur eine geringe Rolle spielen. – (b) Der EPL Wert entspricht dem Durchschnitt aller OECD Indikatoren für die Härte des Beschäftigungsschutzes und die Regulierung von befristeten Arbeitsverträgen. Der Index kann Werte zwischen 0 und 6 annehmen, wobei größere Werte für eine eher strikte Regulierung stehen. Mehr Informationen zu diesem Indikator finden sich in OECD (1999).

Es gibt auch eine Reihe von individuellen Faktoren von denen angenommen wird, dass sie das Arbeitslosigkeitsrisiko beeinflussen. Tabelle 3 stellt die Variablen, die in der Analyse berücksichtigt werden, und ihre Operationalisierung vor. Zur Messung von Beschäftigung und Arbeitslosigkeit wurden die ILO Richtlinien verwendet.⁵⁾ Die Untersuchung verwendet darüber hinaus Dummies für Länder und Zeitperioden. Es wird nicht davon ausgegangen, dass der Einfluss der erklärenden Variablen für Männer und Frauen zu Beginn ihrer Erwerbskarriere signifikant verschieden ist. Aus diesem Grund wird ein gemeinsames Modell spezifiziert und ein Dummy für Männer verwandt.

3. Operationalisierung der individuellen Faktoren	
Variable	Definition
Arbeitsmarkterfahrung und Arbeitsmarkterfahrung quadriert	Zeit seit erstem Eintritt in den primären Arbeitsmarkt in Monaten
Arbeitslosigkeitserfahrung	Dummy: Person war arbeitslos während der letzten fünf Jahre (Referenzgruppe = keine Arbeitslosigkeitserfahrung)
Bildung	Zwei Dummies für hohe Bildung (ISCED 5 – 7) und mittlere Bildung (ISCED 3), (Referenzgruppe = niedrige Bildung, ISCED 0 – 2)
Beschäftigungsverhältnis	Drei Dummies für befristet, geringfügige Teilzeitbeschäftigung, selbstständig (Referenzgruppe = unbefristet)

4 Ergebnisse

Ergebnisse der Zeitreihenanalyse: Zum Einfluss von Marktfaktoren und institutionellen Faktoren auf Jugendarbeitslosigkeit

Die Ergebnisse der Regressionsmodelle mit Zufallseffekten und fixen Effekten für Arbeitslosigkeitsraten sind in Tabelle 4 zusammengefasst. Diese Ergebnisse zeigen, dass alle Markt- und institutionellen Faktoren eine signifikante Rolle spielen. Wie erwartet hängt die Arbeitslosigkeitsrate von der ökonomischen Entwicklung und der Handelsoffenheit eines Landes ab. Arbeitslosigkeitsraten sind geringer in Ländern mit einer hochqualifizierten Erwerbsbevölkerung. Ein geschlossener Arbeitsmarkt ist assoziiert mit hoher Arbeitslosigkeit und Arbeitslosigkeitsraten sind höher in Ländern mit starkem Beschäftigungsschutz. Bildungsanforderung ist allerdings nur in dem Regressionsmodell mit fixen Effekten ein signifikanter Prädiktor. Die Ergebnisse des Breusch und Pagan Lagrangian Multiplier Tests zeigen, dass die zentrale Annahme des Modells mit Zufallseffekten verletzt ist. Das Ergebnis des Hausman Tests verdeutlicht, dass das Modell mit fixen Effekten dem Modell mit Zufallseffekten vorgezogen werden sollte. Es ist jedoch wichtig zu betonen, dass beide Regressionsmodelle zu den gleichen Schlussfolgerungen über die Rolle von Marktfaktoren und institutionellen Faktoren führen.

Ergebnisse der Analyse von Individualdaten: Zum Einfluss von individuellen, Markt- und institutionellen Faktoren auf Übergänge in Arbeitslosigkeit

Im nächsten Schritt werden individuelle Faktoren analysiert. Tabelle 5, Modell 1, macht deutlich, dass Bildung, Beschäftigungsverhältnis und bisherige Erwerbsgeschichte signifikante Prädiktoren des Eintritts in Arbeitslosigkeit sind. Je mehr Arbeitsmarkterfahrung Individuen akkumuliert haben, desto geringer ihr Arbeitslosigkeitsrisiko. Nach einem gewissen Erfahrungszuwachs schwächt sich dieser Effekt allerdings ab. Karriereinstabilität – hier durch den Indikator vorange-

5) Die berufliche Ausbildung wird in dieser Analyse nicht berücksichtigt, da sie noch als Teil der Bildung und nicht als eine erste Beschäftigung angesehen wird. Darüber hinaus kommt der beruflichen Ausbildung in den ausgewählten Ländern eine ganz unterschiedliche Bedeutung für die Integration in den Arbeitsmarkt zu.

**4. Ergebnisse der Regressionsmodelle mit Zufallseffekten und fixen Effekten:
Der Einfluss von Marktfaktoren und institutionellen Faktoren auf Jugendarbeitslosigkeit**

	Regressionsmodell mit			
	Zufallseffekten		fixen Effekten	
	Koeffizient	Standardfehler	Koeffizient	Standardfehler
Wirtschaftliche Entwicklung	-0,009	0,005	-0,030	0,007
Handelsoffenheit	-0,040	0,006	-0,051	0,007
Bildungsanforderung (ln)	-0,155	0,189	-0,312	0,180
Geschlossenheit des Arbeitsmarktes (ln)	0,338	0,041	0,262	0,042
Härte des Beschäftigungsschutzes	0,103	0,028	0,124	0,029
Konstante	-1,739	0,805	-0,802	0,769
Beobachtungen		134		134
Gruppen		20		20
R ² innerhalb		0,732		0,756
R ² zwischen		0,387		0,278
R ² gesamt		0,420		0,306
Sigma_u		0,423		0,810
Sigma_e		0,112		0,112
Rho		0,934		0,981
Breusch and Pagan Test	chi2(1) = 296,4, Prob>chi2 = 0,000			
Hausman Test	chi2(5) = 30,93, Prob>chi2 = 0,000			

Quelle: OECD (2004), eigene Berechnungen.

gangene Phasen der Arbeitslosigkeit erfasst – erhöht das Risiko einer Arbeitslosigkeit. Je höher der Bildungsstand einer Person, desto geringer das Risiko, arbeitslos zu werden. Flexible Beschäftigungsverhältnisse erhöhen hingegen dieses Risiko. Zwischen Männern und Frauen ergeben sich keine signifikanten Unterschiede.

Mit Ausnahme der ökonomischen Entwicklung bleiben alle Marktfaktoren und institutionellen Faktoren signifikante Prädiktoren, wenn in dem Modell individuelle Faktoren kontrolliert werden. Handelsoffenheit ist mit einem geringeren Arbeitslosigkeitsrisiko verbunden. Übergänge in Arbeitslosigkeit sind auch signifikant und negativ verknüpft mit der Geschlossenheit des Arbeitsmarktes. Je stärker der Beschäftigungsschutz, desto größer das Arbeitslosigkeitsrisiko. Die Annahme, dass ein hoher Anteil an hochqualifizierten Arbeitnehmerinnen und Arbeitnehmern die Arbeitslosigkeitsrate senkt, wird hingegen nicht bestätigt. Wenn in dem Modell individuelle Fähigkeiten kontrolliert werden, wird dieser institutionelle Variableneffekt positiv und signifikant. Dieses Ergebnis könnte auf die gewählte Operationalisierung zurückzuführen sein. Die Bildungssysteme der betrachteten Länder unterscheiden sich in mehrfacher Hinsicht und somit variiert auch das Alter mit dem gewisse Bildungszertifikate erworben werden. Es sollte aber hervorgehoben werden, dass dieses Ergebnis repliziert wird, wenn andere Indikatoren, wie beispielsweise der Anteil der Bevölkerung im Alter von 25 bis 34 Jahren mit einem tertiären Abschluss, verwendet werden. Ein weiterer möglicher Erklärungsfaktor ist die ungenaue Messung von Bildung im ECHP, die es nicht ermöglicht, die länderspezifischen Besonderheiten hinreichend abzubilden. Schließlich soll an dieser Stelle auch darauf hingewiesen werden, dass der Einfluss der Marktfaktoren und institutionellen Variablen auf Übergänge in Arbeitslosigkeit in einer sequentiellen Weise untersucht worden ist und die Ergebnisse über alle Spezifikationen robust sind. Aufgrund von Platzgründen werden hier allerdings die Ergebnisse dieser anderen Spezifikationen nicht dargestellt.

Wie verändern sich die Ergebnisse, wenn die Analyse Ländereffekte berücksichtigt? Vergleicht man die Ergebnisse des Modells 1 in Tabelle 5 mit denen des alternativen Ansatzes, der auch Länderdummies berücksichtigt (Modell 2), so kommt man weitestgehend zu den gleichen Schlussfolgerungen. Das Ergebnis, dass Bildung, Erwerbsgeschichte und Beschäftigungsverhältnis eine Rolle spielen, ist robust. Gleiches gilt für die Ergebnisse für die Marktfaktoren und institutionellen Faktoren, wobei die Schätzer auf einem Zehn-Prozent-Niveau signifikant bleiben. Einzige Ausnahme ist der Prädiktor Beschäftigungsschutz, der seine Signifikanz verliert. Diese Veränderungen der Ergebnisse unter Kontrolle der Länderdummies verweist auf die Schwierig-

5. Ergebnisse der Logitmodelle mit Zufallseffekten: Zum Einfluss von individuellen, Markt- und institutionellen Faktoren auf Übergänge in Arbeitslosigkeit				
	Modell 1		Modell 2	
	Koeffizient	Standardfehler	Koeffizient	Standardfehler
Individuelle Faktoren				
Arbeitsmarkterfahrung	-0,034	0,008	-0,034	0,008
Arbeitsmarkterfahrung quadriert	0,000	0,000	0,000	0,000
Arbeitslosigkeitserfahrung	0,497	0,111	0,476	0,113
Bildung				
Hoch	-0,589	0,132	-0,629	0,135
Mittel	-0,416	0,117	-0,399	0,120
Niedrig				
Beschäftigungsverhältnis				
Unbefristet				
Befristet	0,866	0,118	0,832	0,120
Geringfügige Teilzeitbeschäftigung	1,145	0,186	1,131	0,189
Selbstständig	-0,246	0,247	-0,270	0,250
Männer	0,029	0,106	0,026	0,115
Marktfaktoren und institutionelle Faktoren				
Wirtschaftliche Entwicklung	-0,009	0,007	0,022	0,049
Handelsoffenheit	-0,045	0,012	-0,078	0,048
Geschlossenheit des Arbeitsmarktes (ln)	0,285	0,110	0,301	0,154
Bildungsanforderungen (ln)	2,001	0,482	2,049	1,059
Härte des Beschäftigungsschutzes	0,200	0,088	0,094	0,235
Kontrollvariablen				
Zeitperiode				
1995	-1,554	0,779	-1,515	0,835
1996	0,226	0,217	0,209	0,343
1997	0,392	0,201	0,340	0,261
1998	0,536	0,177	0,553	0,200
1999	0,436	0,164	0,444	0,173
2000	0,318	0,156	0,319	0,159
2001				
Länder				
Westdeutschland			-0,667	0,905
Ostdeutschland			-0,602	0,629
Niederlande			1,229	1,485
Frankreich			0,345	0,591
Dänemark			0,596	0,646
Finnland			0,211	0,597
Italien			0,125	0,599
Spanien			1,033	1,542
Portugal			1,759	2,619
Konstante	-11,910	1,993	-12,873	4,843
Anzahl der Beobachtungen	9 688		9 688	
Anzahl der Gruppen	4 133		4 133	
Anzahl der Ereignisse	596		596	

Quelle: ECHP (1994 – 2001), eigene Berechnungen.

keit ländervergleichender Studien, den Effekt einer institutionellen Variable von einem Ländereffekt analytisch zu trennen. Angesichts der Kürze des Untersuchungszeitraumes, der geringen Fallzahl und der starken Pfadabhängigkeiten in den einzelnen Ländern ergibt sich für einige institutionelle Variablen eine zu geringe Variation. Gleichzeitig ist es ein bedeutsames Ergebnis, dass keiner der Länderdummies signifikant ist. Dies deutet darauf hin, dass die Marktfaktoren und institutionellen Faktoren eine gute Approximation der institutionellen Kontexte wiedergeben.

6. Ergebnisse der Logitmodelle mit Zufallseffekten: Zum Einfluss von individuellen, Markt- und institutionellen Faktoren, und Interaktionseffekten auf Übergänge in Arbeitslosigkeit		
	Modell 1	Modell 2
	Koeffizient	Standardfehler
Individuelle Faktoren		
Arbeitsmarkterfahrung	-0,035	0,008
Arbeitsmarkterfahrung quadriert	0,000	0,000
Arbeitslosigkeitserfahrung	0,502	0,109
Bildung		
Hoch	-1,026	0,258
Mittel	-0,428	0,117
Niedrig		
Beschäftigungsverhältnis		
Unbefristet		
Befristet	0,899	0,119
Geringfügige Teilzeitbeschäftigung	1,238	0,187
Selbstständig	-0,256	0,247
Männer	0,129	0,108
Marktfaktoren und institutionelle Faktoren		
Wirtschaftliche Entwicklung	0,003	0,006
Handelsoffenheit	-0,048	0,011
Geschlossener Arbeitsmarkt		
Offener Arbeitsmarkt (erstes Quartal)	-0,318	0,189
Geschlossener Arbeitsmarkt (viertes Quartal)	0,448	0,150
Bildungsanforderung		
Geringqualifizierte Erwerbsbevölkerung (erstes Quartal)	-0,535	0,172
Hochqualifizierte Erwerbsbevölkerung (viertes Quartal)	0,274	0,189
Härte des Beschäftigungsschutzes	0,315	0,106
Interaktionseffekte		
Offener Arbeitsmarkt * hohe Bildung	0,865	0,321
Geschlossener Arbeitsmarkt * hohe Bildung	0,737	0,303
Geringqualifizierte Erwerbsbevölkerung * hohe Bildung	-0,581	0,319
Hochqualifizierte Erwerbsbevölkerung * hohe Bildung	-0,003	0,344
Beschäftigungsschutz * Arbeitslosigkeitserfahrung	-0,298	0,109
Kontrollvariablen		
Zeitperiode		
1995	-1,185	0,787
1996	0,470	0,223
1997	0,683	0,229
1998	0,551	0,177
1999	0,473	0,161
2000	0,356	0,158
2001		
Konstante	-3,270	0,254
Anzahl der Beobachtungen	9 688	
Anzahl der Gruppen	4 133	
Anzahl der Ereignisse	596	

Quelle: ECHP (1994 – 2001), eigene Berechnungen.

Schaut man sich abschließend noch die anderen Ergebnisse in Tabelle 5 an, so zeigt sich, dass die Zeitperiode ein signifikanter Prädiktor des Eintritts in Arbeitslosigkeit ist. Signifikante Unterschiede zwischen Männern und Frauen werden hingegen nicht beobachtet.

Die folgende Analyse schätzt Regressionsmodelle, die Interaktionseffekte berücksichtigen. Um die Interpretation der Ergebnisse zu vereinfachen verwendet dieses Modell Dummyvariablen und unterscheidet eine hoch qualifizierte von einer geringqualifizierten Erwerbsbevölkerung. Dazu werden Quartile verwendet. Auf die gleiche Weise werden auch Dummyvariablen generiert, die zwischen offenen und geschlossenen Beschäftigungssystemen differenzieren. Diese Dummyvariablen und der EPL Wert werden dann mit der Variablen „hoher Bildungsstand“ interagiert. Tabelle 6 zeigt die Ergebnisse des Logitmodells mit Zufallseffekten. Die zuvor beobachteten Unterschiede zwischen Hochqualifizierten und Geringqualifizierten ändern sich. Die Analyse erhärtet die Annahme, dass hochqualifizierte Arbeitnehmerinnen und Arbeitnehmer davon profitieren, wenn sie in einem Land arbeiten, in dem die Erwerbsbevölkerung insgesamt eher geringqualifiziert ist. Der gegenteilige Effekt trifft jedoch nicht zu. Wie erwartet nimmt das Risiko eines Übergangs in Arbeitslosigkeit für junge Erwachsene in offenen und geschlossenen Systemen zu. Arbeitslosigkeitserfahrung erhöht generell das Risiko weiterer Arbeitslosigkeit. Entgegen den Erwartungen ist dieser Effekt jedoch weniger stark in Ländern mit einem ausgeprägten Beschäftigungsschutz.

5 Zusammenfassung und Diskussion

Dieser Beitrag befasst sich mit Jugendarbeitslosigkeit in zehn westeuropäischen Ländern, einem zentralen Indikator für den Übergang von Ausbildung in den Arbeitsmarkt. Ausgangspunkt war die Annahme, dass die sozioökonomischen Veränderungen auf dem Arbeitsmarkt und im Wohlfahrtsstaat Marktrisiken verschärft und den Grad sozialer Sicherung verringert haben. Die Auswirkungen dieser globalen Veränderungen sind jedoch länder- und personenspezifisch.

Die Ergebnisse erhärten die Forschungshypothesen dieses Beitrags. Das erste wichtige Ergebnis war ein Beleg für die Hypothesen über die Rolle der Marktfaktoren und institutionellen Faktoren. Während Handelsoffenheit und hohe Bildungsanforderungen mit niedrigeren Arbeitslosigkeitsraten einhergehen, führt ein geschlossenes Beschäftigungssystem und strikter Beschäftigungsschutz zu mehr Arbeitslosigkeit. Auf der Individualebene zeigt sich, dass Arbeitsmarkterfahrung eine wichtige Quelle von Beschäftigungssicherheit ist, wohingegen flexible Beschäftigung oder vorangegangene Phasen der Arbeitslosigkeit den zukünftigen Erwerbsverlauf eher destabilisieren. Geringqualifizierte Arbeitsmarkteinsteigerinnen und -einsteiger sind einem größeren Arbeitslosigkeitsrisiko ausgesetzt und diejenigen, die in unsicheren Jobpositionen beschäftigt sind, verlieren auch häufiger ihren Job. Die Ergebnisse erhärten darüber hinaus die Annahme, dass die soziale, ökonomische und temporale Unsicherheit für gewisse Gruppen von jungen Arbeitnehmerinnen und Arbeitnehmern in einigen institutionellen Kontexten besonders hoch ist. Geringqualifizierte sind generell häufiger von Arbeitslosigkeit betroffen, aber die Unterschiede zwischen den gering- und hochqualifizierten Arbeitnehmerinnen und Arbeitnehmern sind weniger ausgeprägt in Ländern mit einer hochqualifizierten Erwerbsbevölkerung und in offenen wie auch geschlossenen Beschäftigungssystemen. Vorangegangene Phasen der Arbeitslosigkeit destabilisieren generell die zukünftige Erwerbskarriere, aber dieser Effekt ist weniger stark ausgeprägt in Ländern mit einem starken Beschäftigungsschutz.

Zusammenfassend lässt sich sagen, dass die Länderunterschiede hinsichtlich des Ausmaßes der Jugendarbeitslosigkeit auf Marktfaktoren und institutionelle Unterschiede zurückgeführt werden können und soziale Ungleichheit überproportional in gewissen Gruppen wie zum Beispiel bei Geringqualifizierten und unerfahrenen Arbeiterinnen und Arbeitern mit fragmentierten Karrieren beobachtet wird.

Die empirische Analyse war auf den Zeitraum 1994 – 2001 und zehn westeuropäische Länder beschränkt. Es sind deshalb weiterführende Analysen mit mehr Ländern und längeren Untersuchungszeiträumen wünschenswert. Eine Weiterführung des hier vorgestellten Ansatzes könnte es auch sein, geschlechtsspezifische Effekte und andere Altersgruppen zu analysieren sowie unterschiedliche Modellspezifikationen mit anderen Indikatoren zu entwickeln. Wünschenswert ist auch eine Mehrebenenanalyse, um zu überprüfen, ob die in diesem Beitrag vorgestellten Ergebnisse der Modelle mit Zufallseffekten und fixen Effekten robust sind. Es ist daher meine Hoffnung, dass dieser Beitrag weitere empirische Untersuchungen anregt, die die Effekte von Marktfaktoren, institutionellen und individuellen Faktoren auf Jugendarbeitslosigkeit analysieren.

Literatur

Allison (1982). *Discrete-time Methods for the Analysis of Event Histories*, in Leinhard, S. (Hrsg.) *Sociological Methodology*. San Francisco: Jossey Bass, 61 – 98.

Allison, P. (1995). *Survival analysis using the SAS system. A practical guide*. Cary, North Carolina, USA: SAS Institute Inc.

Allmendinger, J. (1989). *Educational System and Labour Market Outcomes*, *European Sociological Review*, 3, 231 – 250.

Begg, C. and Gray, R. (1984). *Calculation of polychotomous logistic regression parameters using individualized regressions*, *Biometrika*, 71, 1, 11 – 18.

Blossfeld, H.-P. (2002). 'Globalisation, social inequality and the role of country-specific institutions. Open research questions in a learning society', in P. Conceição, Heitor, H.V. und Lundvall, B.-A. (Hrsg.) *Innovation, Competence Building and Social Cohesion in Europe: Towards a Learning Society*, Cheltenham/Lyme: Edward Elgar.

Blossfeld, H.-P., Klijzing, E., Mills, M. und Kurz, K. (Hrsg.) (2005) *The losers in a globalizing world: becoming an adult in uncertain times*. London/New York: Routledge.

Blossfeld, H.-P. and Stockmann, R. (1998/1999). *The German Dual System in Comparative Perspective*, *International Journal of Sociology*, 28, 4, 3 – 28.

Breen, R. (2005). *Explaining Cross-national Variation in Youth Unemployment. Market and Institutional Factors*, *European Sociological Review*, 21, 2, 125 – 134.

Ebbinghaus, B. (2005). When less is more. Selection Problems in Large-N and Small-N Cross-National Comparisons, *International Sociology*, 20, 2, 133 – 152.

European Commission Eurostat (1996). *The European Community Household Panel (ECHP): Survey Methodology and Implementation*, Vol. 1, Luxembourg: Eurostat.

Esping-Andersen, G. (1993) (Hrsg.). *Changing Classes. Stratification and Mobility in Post-industrial Societies*. London/Newbury Park/New Delhi: Sage Publications.

Esping-Andersen, G. (1996). *Welfare states in transition. National adaptations in global economies*. London/ Thousands Oaks/New Delhi: Sage Publications.

Esping-Andersen, G. (1999). *Social foundations of postindustrial economics*. Oxford: Oxford University Press.

Esping-Andersen, G. und Regini, M. (2000). *Why Deregulate Labour Markets?* Oxford: Oxford University Press.

Gangl, M. (2001). European patterns of labour market entry. A dichotomy of occupationalized vs. non-occupationalized systems? *European Societies*, 3, 4, 471 – 494.

Gangl, M. (2003). The only way is up? Employment Protection and job mobility among recent entrants to European labour markets, *European Sociological Review*, 19, 5, 429 – 449.

Golsch, K. (2005). *The Impact of Labour Market Insecurity on the Work and Family Life of Men and Women*, Europäische Hochschulschriften, Reihe 22: Soziologie, Frankfurt a.M.: Lang.

Heery, E. und Salmon, J. (Hrsg.) (2000). *The Insecure Workforce*. London/ New York: Routledge.

Kittel, B. (2005). Pooled Analysis in der ländervergleichenden Forschung: Probleme und Potenziale, in S. Kropp und M. Minkenberg (Hrsg.) *Vergleichen in der Politikwissenschaft*. Wiesbaden: Westdeutscher Verlag.

Ladipo, D. und Wilkinson, F. (2002). 'More pressure, less protection', in Burchell, B., Ladipo, D. und Wilkinson, F. (Hrsg.) *Job Insecurity and Work Intensification*. London/New York: Routledge.

Mayer, K.-U. (1997). Notes on the comparative political economy of life courses, *Comparative Social Research*, 16, 203 – 226.

Mills, M. und Blossfeld, H.-P- (2005). Globalization, Uncertainty and the Early Life Course. A Theoretical Framework, in Blossfeld, H.- P., Klijzing, E., Mills, M. und Kurz, K. (Hrsg.). *Globalization, Uncertainty, and Youth in Society*. London/ New York: Routledge.

OECD (1999). *Employment outlook*. Paris: OECD.

Regini, M. (1997). Different Responses to Common Demands: Firms, Institutions, and Training in Europe, *European Sociological Review*, 13, 267 – 282.

Regini, M. (2000). Between deregulation and social pacts: the responses of European economies to globalization, *Politics and Society*, 28, 1, 5 – 33.

Russell, H. und O'Connell, P. (2001). Getting a Job in Europe: The transition from Unemployment to work among Young People in Nine European Countries, *Work, Employment & Society*, 15, 1, 1 – 24.

Shavit, Y. und Müller, W. (Hrsg.) (1998). *From School to Work. A Comparative Study of Educational Qualifications and Occupational Destinations*. Oxford: Clarendon Press.

Wooldridge, J. (2003). *Introductory econometrics: a modern approach*. South Western College Publishing.

Autoren- und Kontaktangaben:

Dr. Katrin Golsch, Wirtschafts- und Sozialwissenschaftliche Fakultät Universität zu Köln,
Herbert-Lewin-Str. 2, 50931 Köln, E-Mail: golsch@wiso.uni-koeln.de

Möglichkeiten und Grenzen der Nutzung der Einzeldaten der amtlichen Kinder- und Jugendhilfestatistik

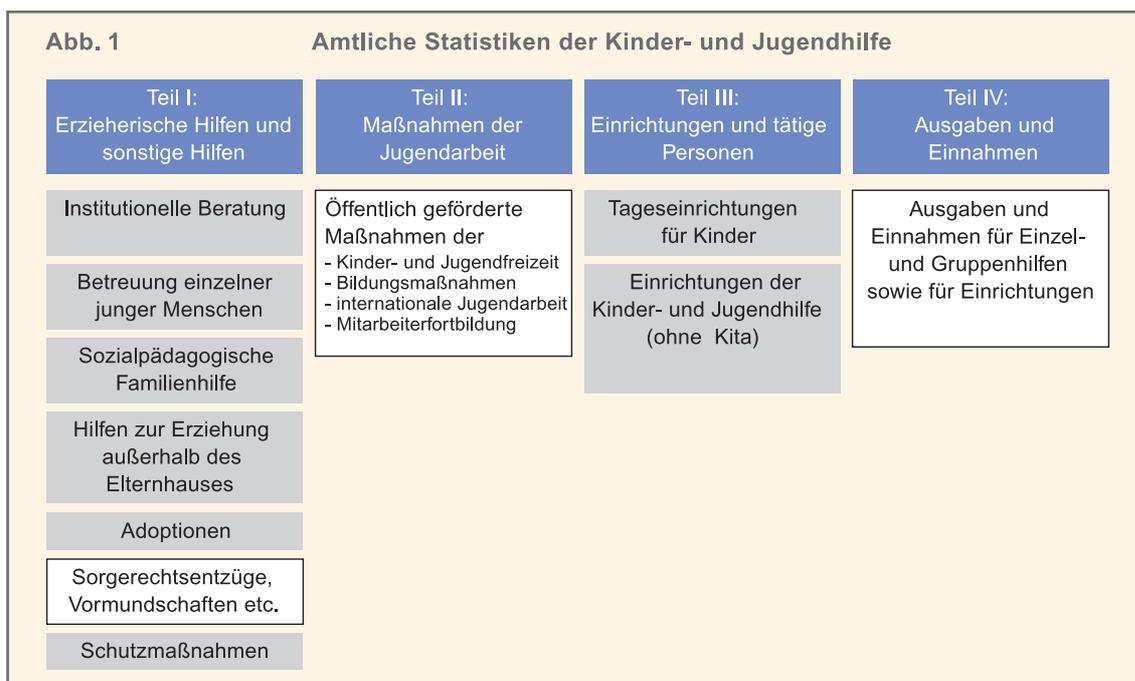
Ein Erfahrungsbericht

Matthias Schilling

Mehre amtliche Statistiken zur Kinder- und Jugendhilfe basieren auf personenbezogenen Einzeldaten, die sich für die wissenschaftliche Analyse besonders gut eignen. Die Dortmunder Arbeitsstelle Kinder- und Jugendhilfestatistik nutzt diese Mikrodaten bereits seit mehreren Jahren, wie in einigen exemplarischen Auswertungen aufgezeigt wird. Allerdings gestaltete sich die bisherige Beschaffung der Daten äußerst aufwändig, da Verabredungen mit 16 statistischen Landesämtern und 16 Datenschutzbeauftragten getroffen werden mussten. Durch das Angebot des Forschungsdatenzentrums wird der Zugang erheblich erleichtert, sodass die Arbeitsstelle die Mikrodaten intensiver nutzen wird.

Die amtliche Kinder- und Jugendhilfestatistik existiert zwar seit den Anfängen des letzten Jahrhunderts, aber insbesondere mit der Einführung des Sozialgesetzbuches VIII, Kinder- und Jugendhilfegesetz (SGB VIII) Anfang der 90er Jahre wurde der Erhebungsumfang insbesondere bei den Hilfen zur Erziehung deutlich ausgeweitet. Bei der Konzeption der neuen Statistiken wurde methodisch erstmals in den Sozialstatistiken die Individualdatenerfassung eingeführt (vgl. Hoffmann 1991). Einige Jahre später wurde dieses Prinzip auch bei der Sozialhilfestatistik angewendet.

Die amtliche Kinder- und Jugendhilfestatistik umfasst inzwischen mehrere einzelne Statistiken, die vier Teilbereichen zugeordnet werden. Dies sind Statistiken zu den Hilfen zur Erziehung und anderen Hilfen, Maßnahmen der Kinder- und Jugendarbeit, Einrichtungs- und Personal-



daten sowie Angaben zu den Ausgaben und Einnahmen für die Kinder- und Jugendhilfe (vgl. Abbildung 1; ausführlich Rauschenbach/Schilling 1997; Schilling 2002). Grau unterlegt sind in der Abbildung 1 diejenigen Statistiken, die auf Individualdatensätzen beruhen und deshalb für die wissenschaftliche Analyse besonders geeignet sind. Bei den Einrichtungen und tätigen Personen beziehen sich die Einzeldatensätze auf Einrichtungen bzw. auf die tätigen Personen.

Im Folgenden werden die Erfahrungen der Dortmunder Arbeitsstelle Kinder- und Jugendhilfestatistik mit den Analysen der Einzeldaten vorgestellt. Die Arbeitsstelle existiert seit 1997 an der Universität Dortmund und beschäftigt sich mit der wissenschaftlichen Analyse der Erhebungsergebnisse der amtlichen Kinder- und Jugendhilfestatistik. Ziel der Arbeitsstelle ist es, die umfangreichen empirischen Daten zur Kinder- und Jugendhilfe aufzubereiten, auszuwerten und in den Fachdiskurs der Kinder- und Jugendhilfe einzuspeisen (www.akj-stat.fb12.uni-dortmund.de).

Generell werden für die Analysen der Arbeitsstelle die Ergebnistabellen der statistischen Ämter des Bundes und der Länder herangezogen. Mit den Standardtabellen können ca. 80 % der Forschungsfragen bearbeitet werden. Für die verbleibenden 20 % wären Merkmalskombinationen notwendig, die in der Vorrattabellierung der statistischen Ämter nicht vorgesehen sind und aus ökonomischen Gesichtspunkten auch nicht vorsehbar sind. Deshalb hatte die Arbeitsstelle im Rahmen der wissenschaftlichen Nutzung der amtlichen Einzeldaten gemäß § 16 Abs. 6 BStG für gesonderte Forschungsfragen bei den einzelnen statistischen Landesämtern immer wieder anonymisierte Einzeldaten angefordert. Dabei hat sich allerdings herausgestellt, dass die Beschaffung der Einzeldatensätze, insbesondere wenn es sich um Forschungsfragen handelt, die sich auf ganz Deutschland beziehen, äußerst aufwändig ist, da mit 16 Landesämtern und mit 16 Datenschutzbeauftragten kooperiert werden musste. Deshalb kann die Bereitstellung aller Landesergebnisse durch ein Forschungsdatenzentrum nur als erhebliche Organisationsvereinfachung gelobt werden.

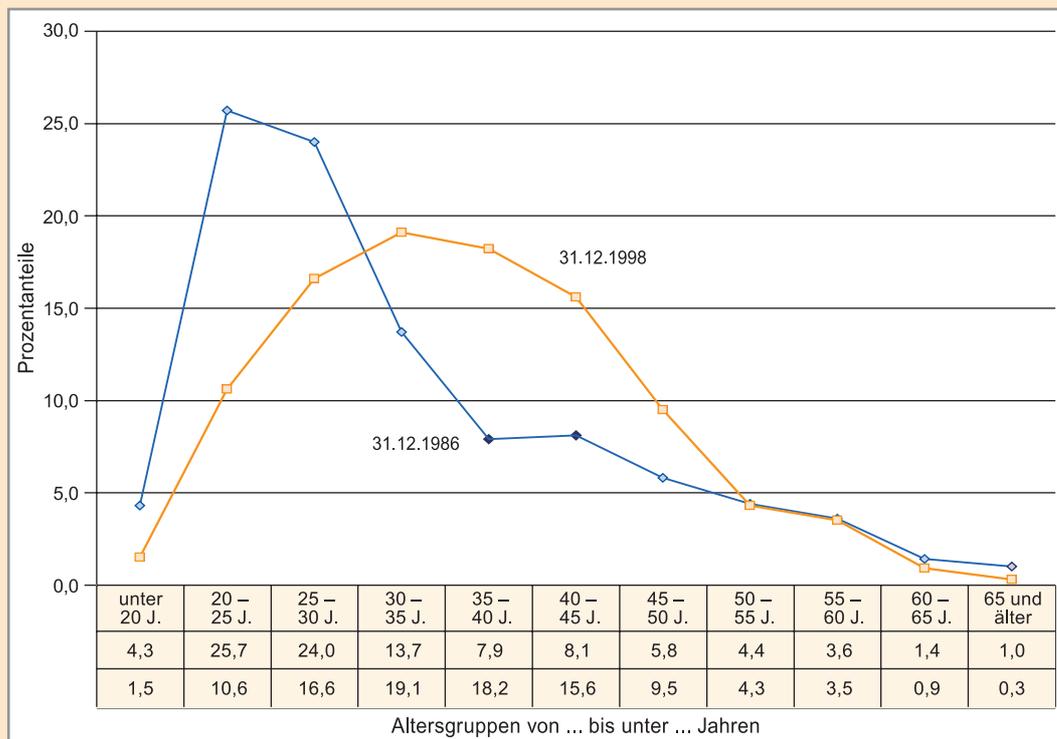
1 Bisherige Nutzungskontexte

Bereits zu Beginn der Arbeitsstelle Kinder- und Jugendhilfestatistik wurden exemplarisch die Einzeldaten der Hilfen zur Erziehung außerhalb des Elternhauses (§§ 32 – 35 SGB VIII) in Nordrhein-Westfalen einer ausführlichen Analyse unterzogen (vgl. Biermann 1997). Neben mehreren inhaltlich interessanten Ergebnissen wurde von Biermann (1997) der Mehrnutzen der Einzeldaten herausgestellt:

- einfachere individuelle Zusammenstellung der notwendigen Analysetabellen,
- Erstellung und Analyse von Merkmalskombinationen, die nicht in den Standardtabellen vorgesehen sind und
- Bildung von Subgruppen und deren statistischer Analyse.

Eine weitere Analyse der Einzeldaten wurde zum Bereich der Fachkräfte in den Erziehungshilfen durchgeführt. Mit den Standardtabellen kann nur nachgewiesen werden, dass die älteren Mitarbeiter/-innen im Gruppendienst zwischen 1982 und 1998 deutlich zugenommen haben. So sind z. B. die Anteile der Altersgruppen der 35- bis unter 45-Jährigen von 8 % auf 18 % bzw. 16 % gestiegen (vgl. Abbildung 2).

Abb. 2 Tätige Personen im Gruppendienst in Einrichtungen der Heimerziehung nach Alter (alte Bundesländer; 1982, 1998; %-Anteile je Stichtag)



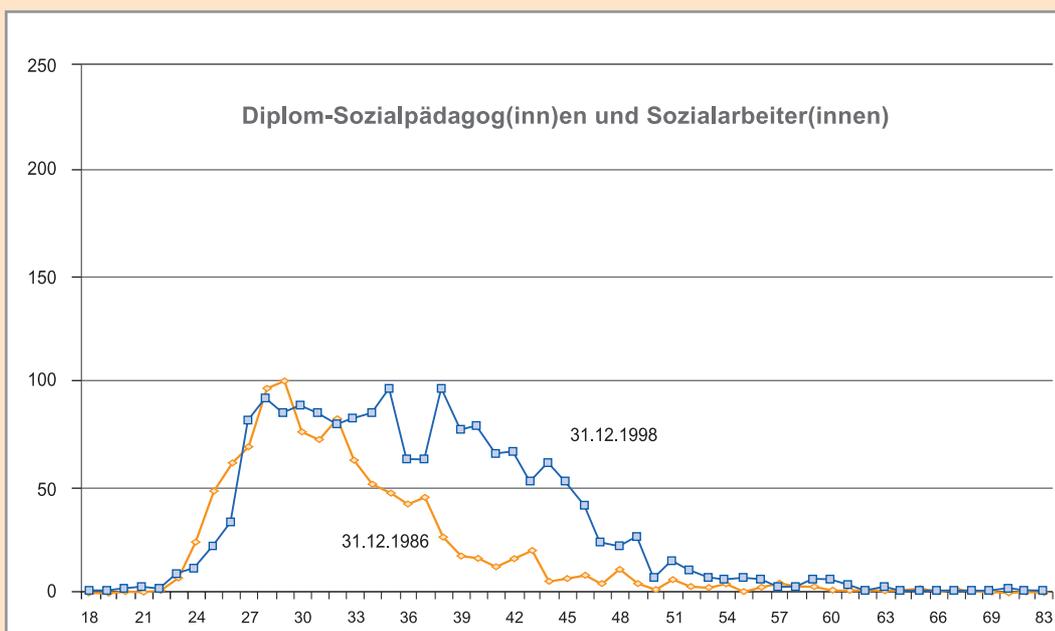
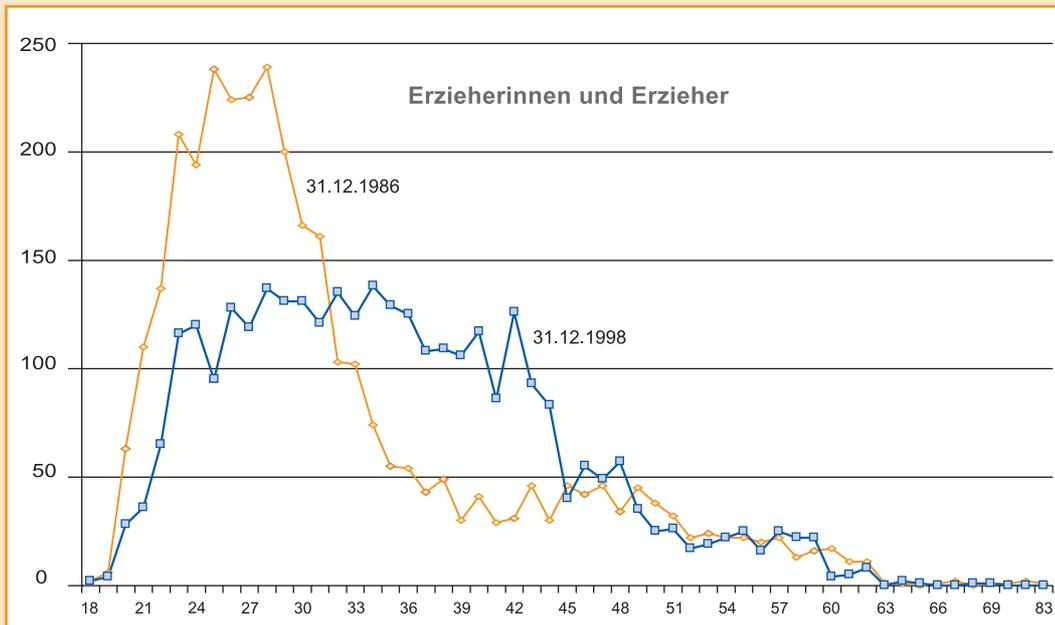
Quelle: Statistisches Bundesamt: Statistiken der Kinder- und Jugendhilfe – Einrichtungen und tätige Personen, verschiedene Jahrgänge, Wiesbaden

Darüber hinaus ist allerdings auch von Interesse, ob sich in der Altersstruktur und damit in der Berufsbiographie zwischen den wichtigsten Berufsgruppen in diesem Tätigkeitsfeld, den Erzieher(inne)n und Diplom-Sozialpädagog(inn)en Unterschiede zeigen. Für die beiden Berufsgruppen zeigen sich deutliche Verschiebungen zu den älteren Jahrgängen, allerdings mit unterschiedlichen Ausprägungen (vgl. Abbildung 3).

Für die Berufsgruppe der Erzieher/-innen ergibt sich, dass sich die Gruppe der jüngeren Erzieher/-innen zwischen 22 und 30 Jahren offensichtlich innerhalb von 10 Jahren halbiert hat.¹⁾ Betrachtet man in analoger Weise die Veränderungen bei den Diplom-Sozialpädagog(inn)en zeigt sich im gleichen Zeitraum für die unter 30-Jährigen kaum eine Veränderung. Somit bleiben die Diplom-Sozialpädagog(inn)en offensichtlich länger in diesem Arbeitsfeld tätig als die Erzieher/-innen. Ebenso zeigt sich, dass die Anzahl der jüngeren Diplom-Sozialpädagog(inn)en, also die Berufseinsteiger, über die Jahre relativ konstant geblieben ist, wohingegen die Anzahl der Berufseinsteiger bei den Erzieher(inne)n deutlich abgenommen hat. In Kombination mit der geringeren Ausstiegsquote bei den Diplom-Sozialpädagog(inn)en ist zu erwarten, dass diese Berufsgruppe kontinuierlich wachsen wird und die der Erzieher/-innen leicht abnehmen wird. Wenn sich diese Tendenz weiter fortsetzt, was aufgrund des steigenden Anforderungsprofils zu erwarten ist, wird die Gruppe der Diplom-Sozialpädagog(inn)en in den nächsten Jahren weiter anwachsen und zumindest bei den über 35-Jährigen ein annähernd ausgeglichenes Verhältnis zu den Erzieher/-innen erreicht haben (vgl. Schilling 2001).

1) Da die Statistik keine biographischen Verläufe abbildet, sondern nur Häufigkeitsverteilungen zwischen zwei Erhebungszeitpunkten erfasst, kann zwar nicht auf den Verbleib einzelner Personen geschlossen werden, allerdings lässt die Verschiebung der Häufigkeitsverteilung mit einem ähnlichen Muster vermuten, dass es sich wahrscheinlich zu einem Großteil um dieselben Personen handelt.

Abb. 3 Erzieher(innen) und Diplom-Sozialpädagog(inn)en im Gruppendienst in Einrichtungen der Heimerziehung nach Alter (NRW; 1986,1998)



Quelle: Landesamt für Datenverarbeitung und Statistik NRW: Statistiken der Kinder- und Jugendhilfe – Einrichtungen und tätige Personen, verschiedene Jahrgänge; Sonderauswertung der anonymisierten Einzeldaten durch die Arbeitsstelle Kinder- und Jugendhilfestatistik

Ein zentraler Verwendungskontext der Einzeldaten der amtlichen Kinder- und Jugendhilfestatistik hat sich für die Arbeitsstelle in der Analyse der Hilfen zur Erziehung (§§ 27 – 35, 41 SGB VIII) in Nordrhein-Westfalen ergeben. Jährlich wird ein ausführlicher Bericht zur Entwicklung und aktuellen Situation der Hilfen zur Erziehung (HzE) in NRW erstellt.²⁾ Die Zielgruppe des

²⁾ Die Berichte der letzten drei Jahre: Schilling/Pothmann/Overmann (2003); Schilling/Pothmann/Overmann (2004); Schilling/Pothmann/Fischer (2005).

HZE-Berichtes NRW sind Fachleute in den Jugendämtern. Daher war es von Anfang an wichtig, dass in den Analysen die Ergebnisse auch bis auf die Ebene der Jugendämter herunter gebrochen werden. Um das gesamte Analysepotenzial auf der Ebene der Jugendämter ausnutzen zu können, muss auf die Einzeldaten zurückgegriffen werden. Aufgrund der notwendigen Anonymisierung konnte allerdings nicht die Regionalkennung im Datensatz beibehalten werden. Deshalb wurde unter fachlichen Gesichtspunkten eine Typisierung der Jugendämter nach Größenklassen und sozialen Belastungssituationen gewählt. Aus dieser Matrix ergaben sich 16 Jugendamtstypen, die aufgrund der ungleichen Besetzung auf neun Jugendamtstypen zusammengefasst werden mussten (vgl. Tabelle 1). Somit wurde jedes Jugendamt einem bestimmten Typus zugeordnet. Der Typus mit der Ausprägung 1 bis 9 wurde dann anstatt der Regionalkennung vom Landesamt für Datenverarbeitung und Statistik (LDS) NRW in die Einzeldatensätze eingefügt. Mit diesem anonymisierten Einzeldatensatz konnten dann die Auswertungen vorgenommen werden.

1. Übersicht zur Verteilung der Jugendämter in Nordrhein-Westfalen über die Belastungsklassen und Strukturtypen (modifiziertes Ergebnis)					
	kreisfreie Stadt	Landkreis	kreisangehörige Stadt/Gemeinde mit unter 60 000 Einw.	kreisangehörige Stadt/Gemeinde mit 60 000 und mehr Einw.	Insgesamt
Belastungstyp 1	11		13	19	11
Belastungstyp 2	12		36	16	44
Belastungstyp 3		6	37		58
Belastungstyp 4		23			60
Insgesamt	23	29	86	35	173

Quelle: Schilling/Pothmann/Overmann (2003)

Ein weiterer wichtiger Punkt für die Analysen der Arbeitsstelle im Rahmen des HZE-Berichtes ist, dass nicht bei allen Auswertungen auf die Zuordnung zu den einzelnen Jugendämtern verzichtet werden kann. Die einzelnen Jugendämter sollten sich ja im HZE-Bericht mit ihrer eigenen Leistungsbilanz wiederfinden.

Hierfür wurde mit dem LDS NRW ein anderes Anonymisierungsverfahren entwickelt. Datenschutzrechtlich muss nicht der Name des Jugendamtes geschützt werden, sondern die personenbezogenen Angaben der Ergebnistabellen dürfen nicht zu einer Reidentifizierung des einzelnen Hilfeempfängers führen. Deshalb wurden von der Arbeitsstelle fachlich begründete zentrale Eckwerte zu den Hilfen zur Erziehung entwickelt. Diese Eckwerte wurden dann vom LDS NRW zusammengestellt, berechnet und in Tabellenform ausgegeben (vgl. Tabelle 2). Um eine Reidentifizierung auszuschließen, wurden dann, in üblicher Weise, Tabellenfelder mit Angaben kleiner drei geschwärzt. Der Nachteil dieses Verfahrens ist, dass die Arbeit ausschließlich von den statistischen Landesämtern geleistet werden kann und dies bei den knappen Personalressourcen nicht immer möglich ist.

Durch diese Beispiele soll deutlich werden, dass die Analysen mit den Einzeldaten in der Kinder- und Jugendhilfe bisher zu wichtigen Ergebnissen geführt haben. Der entscheidende Nachteil bei der bisherigen Vorgehensweise der Arbeitsstelle war, dass die zur Verfügung stehenden Einzeldaten aus organisatorischen und ökonomischen Gründen regional auf einzelne Bundesländer begrenzt waren. Der Aufwand für den kontinuierlichen Kontakt mit 16 Bundesländern war einfach zu hoch.

2. Stationäre Hilfen in den Jugendämtern Nordrhein-Westfalens nach Alter bei Beginn, Dauer und Abschluss der Maßnahme; 2001 (Jugendämter sind innerhalb der Jugendamtstypen alphabetisch sortiert)							
Jugend- amts- typ	Verwaltungsbezirk	Begonnene Leistungen nach § 34 SGB VIII					
		Anzahl der beg. Hilfen für Minderj. (abs.)	Angaben relativiert auf die altersentsprechende Bevölkerung				
			Insgesamt unter 18 Jahre	unter 6 Jahre	6 bis unter 12 Jahre	12 bis unter 15 Jahre	15 bis unter 18 Jahre
1	Aachen, krfr. Stadt	43	10,8	4,6	7,5	13,1	26,8
1	Dortmund, krfr. Stadt	121	11,7	1,6	3,7	24,0	33,1
1	Duisburg, krfr. Stadt	330	35,5	13,4	28,3	53,6	70,0
1	Düsseldorf, krfr. Stadt	191	21,9	5,7	16,4	33,6	54,6
1	Gelsenkirchen, krfr. Stadt	68	13,4	1,3	9,4	33,2	20,3
1	Herne, krfr. Stadt	56	18,3	1,1	9,8	37,4	44,5
1	Köln, krfr. Stadt	450	27,8	11,6	20,4	50,7	53,3
1	Krefeld, krfr. Stadt	74	17,1	3,8	14,4	23,9	37,8
2	Bielefeld, krfr. Stadt	91	15,4	6,1	5,7	26,1	41,8
2	Bochum, krfr. Stadt	158	25,5	6,9	18,0	47,1	49,2
2	Bonn, krfr. Stadt	88	16,4	3,8	8,0	34,7	40,9
2	Bottrop, krfr. Stadt	71	31,5	15,2	14,3	49,3	72,2
2	Essen, krfr. Stadt	208	21,2	5,4	13,2	34,2	50,2
2	Hagen, krfr. Stadt	58	15,3	4,4	11,6	25,7	31,0

Quelle: Schilling/Pothmann/Overmann (2003)

2 Perspektiven der Nutzung

Durch das Angebot des Forschungsdatenzentrums der Statistischen Landesämter wird die Nutzung der Einzeldaten in der Arbeitsstelle Kinder- und Jugendhilfestatistik erheblich ausgeweitet werden. Bisher stehen die gesamten Ergebnisse für die Statistik »Einrichtungen und tätige Personen in der Kinder- und Jugendhilfe« für die Stichtage 31. 12. 1998 und 31. 12. 2002 zur Verfügung. Die Arbeitsstelle hat bereits begonnen diese Einzeldaten nicht nur für bivariate Auswertungen, sondern auch für multivariate Auswertungen zu nutzen. Fragestellungen sind dabei z. B.: Welches sind die entscheidenden Einflussgrößen für befristete Beschäftigungsverhältnisse in der Kinder- und Jugendhilfe? In welchen Bereichen bestehen die größten Chancen für die Beschäftigung hochschulausgebildeter Fachkräfte?

Aufgrund der positiven Erfahrungen hat die Arbeitsstelle alle weiteren Einzeldaten der amtlichen Kinder- und Jugendhilfestatistik angefordert. Zurzeit wird hierfür ein Anonymisierungskonzept erarbeitet.

Durch die Vorarbeiten der Arbeitsstelle Kinder- und Jugendhilfestatistik stehen anderen wissenschaftlichen Einrichtungen die Einzeldaten prinzipiell auf Antrag auch zur Verfügung. Hier ist zu hoffen, dass diese neue Chance dann auch von anderen Hochschulen genutzt wird.

Literatur

Biermann, B.: Abschluss oder Abbruch: Wie endet die Fremdplatzierung? Möglichkeiten der Auswertung von Einzeldaten, in: Th. Rauschenbach, M. Schilling (Hrsg.), Die Kinder- und Jugendhilfe und ihre Statistik. Band II: Analysen, Befunde und Perspektiven, Neuwied 1997, S. 87 – 114.

Hoffmann, U.: Neuordnung der Jugendhilfestatistik, in: Wirtschaft und Statistik, 43. Jg., 1991, Heft 3, S. 153 – 164.

Rauschenbach, Th./Schilling, M.: Die Kinder- und Jugendhilfe und ihre Statistik. Band 1: Einführung und Grundlagen, Neuwied u. a. 1997.

Schilling, M.: Die Fachkräfte in den Erziehungshilfen, in: V. Birtsch, K. Münstermann, W. Trede (Hrsg.), Handbuch Erziehungshilfen. Leitfaden für Ausbildung, Praxis und Forschung, Münster 2001, S. 458 – 487.

Schilling, M.: Die amtliche Kinder- und Jugendhilfestatistik. Dissertation am Fachbereich Erziehungswissenschaft und Soziologie der Universität Dortmund, Dortmund (<https://eldorado.uni-dortmund.de/handle/2003/2907>) 2002.

Schilling, M./Pothmann, J./Fischer, J. (Hrsg.): Inanspruchnahme und Gewährung von Hilfen zur Erziehung in Nordrhein-Westfalen. HzE Bericht 2003, Dortmund und Köln/Münster 2005.

Schilling, M./Pothmann, J./Overmann, R.: Erziehungshilfen in Nordrhein-Westfalen. HzE Bericht 2001, Dortmund 2003.

Schilling, M./Pothmann, J./Overmann, R.: Hilfen zur Erziehung in Nordrhein-Westfalen. HzE Bericht 2002, Dortmund u. a. 2004.

Autoren- und Kontaktangaben:

Dr. Matthias Schilling, Leiter der Dortmunder Arbeitsstelle Kinder- und Jugendhilfestatistik an der Universität Dortmund, E-Mail: schilling@fb12.uni-dortmund.de

Analyse privater Vermögen auf der Basis der Einkommens- und Verbrauchsstichproben

Peter Westerheide

Der Beitrag beschreibt wesentliche Charakteristika der Einkommens- und Verbrauchsstichproben (EVS) des Statistischen Bundesamtes. Im Mittelpunkt der Betrachtungen stehen die Möglichkeiten, mittels der EVS die Vermögenshöhe, Vermögensstruktur und Vermögensverteilung im Sektor der privaten Haushalte zu analysieren. Thematisiert werden insbesondere Probleme des zeitlichen Vergleichs der EVS-Vermögensdaten. Darüber hinaus werden die Vermögensdaten der EVS mit anderen Datenquellen (Finanzierungsrechnung der Deutschen Bundesbank, Sozio-oekonomisches Panel) verglichen.

Abschließend präsentiert der Beitrag einige Ergebnisse der jüngsten EVS-Erhebung 2003 zu den Geld- und Immobilienvermögen der privaten Haushalte. Danach verfügten die Haushalte im Durchschnitt über ein materielles Nettogesamtvermögen (definiert als Summe von Immobilienvermögen und Geldvermögen) von rund 133 000 Euro. Der Median des Nettovermögens lag mit 50 000 Euro allerdings erheblich niedriger. Es zeigen sich erhebliche Unterschiede zwischen West- und Ostdeutschland. Während die westdeutschen privaten Haushalte im Durchschnitt über ein Geld- und Immobilienvermögen von rund 149 000 Euro verfügen, umfassen die Vermögen der ostdeutschen Haushalte mit knapp 60 000 Euro im Durchschnitt nur 40 % des Vermögens der westdeutschen Haushalte. Eine Analyse der Vermögensverteilung für Deutschland zeigt eine erhebliche Ungleichverteilung des privaten Vermögens. Die Haushalte in der unteren Hälfte der Vermögensverteilung verfügen über etwas weniger als 4 % des gesamten Nettovermögens, während die 20 % vermögensstärksten Haushalte rund zwei Drittel des gesamten privaten Geld- und Immobilienvermögens auf sich vereinen.

1 Einleitung

Im Folgenden werden wesentliche Merkmale der Einkommens- und Verbrauchsstichproben (EVS) des Statistischen Bundesamtes beschrieben. Besonderes Gewicht erhalten dabei die Daten zur Vermögenshöhe der privaten Haushalte und die Möglichkeiten, mit diesen Daten die Vermögensstruktur und Vermögensverteilung im Sektor der privaten Haushalte zu beschreiben. In diesem Zusammenhang werden auch weitere Merkmale der Erhebung, insbesondere die Angaben zu den Einkommen, angesprochen. Ein Schwerpunkt wird hier auf die Arbeit mit anonymisierten Mikrodatenfiles (Scientific-Use-Files) gelegt.

Die Einkommens- und Verbrauchsstichproben werden seit 1962/63 in fünfjährigem Rhythmus erhoben. Sie stellen die derzeit breiteste Erhebung in Deutschland zu Einkommen, Ausgaben und Vermögen der privaten Haushalte dar. In der jüngsten Erhebung 2003 waren (auf der Basis der dem ZEW vorliegenden anonymisierten 98 Prozent-Teilstichprobe) 57 143 Haushalte enthalten, davon 45 772 in den alten Bundesländern (inkl. Westberlin) und 11 371 in den neuen Bundesländern (inkl. Ostberlin).

Die EVS ist als Quotenstichprobe organisiert, sie stellt keine Zufallsstichprobe dar. Abhängig von den Merkmalen Bundesland, Haushaltstyp, soziale Stellung des Haupteinkommensbezieher und Höhe des Nettoeinkommens werden auf der Basis des Mikrozensus Erhebungsquoten festgelegt. Die Hochrechnung der Ergebnisse erfolgt ebenfalls auf der Basis dieser Merkmale anhand der Ergebnisse des Mikrozensus.

Eine regionale Differenzierung der Quotenvorgaben und der tatsächlich beteiligten Haushalte zeigt erhebliche Unterschiede in der Erfüllung des Erhebungssolls (Tabelle 1): Im Durchschnitt wurde das Erhebungssoll nur zu rund 70 % erfüllt, allerdings stellen insbesondere Berlin (11,7 %), Niedersachsen (55,8 %) und Hamburg (58 %) Ausreißer nach unten dar, während die Quoten beispielsweise in Mecklenburg-Vorpommern, Rheinland-Pfalz und Sachsen überdurchschnittlich gut erfüllt werden konnten.

1. Buchführende Haushalte in der EVS 1. Halbjahr 2003			
Bundesland	Quotenvorgabe	beteiligte Haushalte	Relation
Deutschland	37 298	26 207	70,3 %
Baden-Württemberg	4 440	3 351	75,5 %
Bayern	5 072	3 960	78,1 %
Berlin	1 891	221	11,7 %
Brandenburg	1 269	987	77,8 %
Bremen	439	325	74,0 %
Hamburg	1 014	588	58,0 %
Hessen	2 756	2 025	73,5 %
Mecklenburg-Vorpommern	925	741	80,1 %
Niedersachsen	3 519	1 964	55,8 %
Nordrhein-Westfalen	7 413	5 859	79,0 %
Rheinland-Pfalz	1 891	1 521	80,4 %
Saarland	601	388	64,6 %
Sachsen	2 157	1 660	77,0 %
Sachsen-Anhalt	1 328	874	65,8 %
Schleswig Holstein	1 382	1 006	72,8 %
Thüringen	1 201	737	61,4 %

Quelle: Statistisches Bundesamt.

Neben anderen Gründen dürften bei der regional unterschiedlichen Erfüllung der Quoten auch die grundsätzlichen Repräsentationsprobleme der EVS eine Rolle gespielt haben: So sind Haushalte mit besonders niedrigen Einkommen und besonders hohen Einkommen in der EVS tendenziell unterrepräsentiert. Besonders hohe Einkommen werden bereits bei der Festlegung der Erhebungsquoten ausgegrenzt: 2003 lag die so genannte Abschneidegrenze bei 18 000 Euro, in den Vorerhebungen 1993 und 1998 bei 35 000 DM monatlichem Haushaltsnettoeinkommen. Darüber hinaus sind Ausländerhaushalte – die z. B. in Berlin und Hamburg einen deutlich höheren Anteil an der Gesamtbevölkerung als in anderen Bundesländern umfassen – in der EVS tendenziell unterrepräsentiert.¹⁾

Die Einkommens- und Verbrauchsstichprobe wird in mehreren Teilen erhoben: Informationen über die Haushaltszusammensetzung, die Ausstattung der Haushalte mit ausgewählten langlebigen Gebrauchsgütern und ihre Wohnverhältnisse sowie die Vermögensverhältnisse werden bereits im so genannten Einführungsinterview erhoben, das zu Beginn der Erhebung durchge-

1) Vgl. zur Problematik der Unterrepräsentierung von Ausländerhaushalten in der EVS Becker et al. (2003), S. 63 und die dort genannte Literatur., vgl. zum Anteil von Ausländern an der Gesamtbevölkerung nach Bundesländern die Bevölkerungsfortschreibung der Statistischen Ämter des Bundes und der Ländern (http://www.statistik-portal.de/Statistik-Portal/de_jb01_jahrtab2.asp, Abrufdatum 07. 10. 2005).

führt wird. Der entsprechende Stichtag in der letzten Welle 2003 war der 1. Januar 2003. Im Einführungsinterview werden die Haushalte auch gebeten, ihr Haushaltsnettoeinkommen in vorgegebene Einkommensklassen einzuordnen. Da Einkommensangaben für viele sozioökonomische Analysen unerlässlich sind, stellen diese Daten – auch wenn sie von den später erhobenen detaillierteren Angaben üblicherweise abweichen – eine wichtige Vorinformation dar, die mit einem erheblichen zeitlichen Vorlauf vor den Detailangaben verfügbar ist.²⁾ In der Praxis werden für einkommensbezogene Analysen mit diesen vorläufigen Angaben häufig Mittelwerte aus den vorgegebenen Einkommensklassen verwendet, um einzelnen Haushalten ein bestimmtes vorläufiges Nettoeinkommen zuzuordnen.

Die detaillierten Angaben der Haushalte zu ihren Einnahmen und Ausgaben werden seit 1998 im Laufe des Erhebungsjahres in einem Rotationsverfahren erhoben. Dabei führt jeweils ein Viertel der Haushalte in der Gesamtstichprobe je Quartal ein so genanntes Haushaltsbuch, in dem alle Einnahmen und alle zum privaten Verbrauch zählenden Ausgaben sowie die Ausgaben zur Bildung von Vermögen eingetragen werden. Zusätzlich wird von einem Fünftel der Stichprobe ein Feinanschreibungsheft geführt, in dem die Ausgaben für Nahrungsmittel, Getränke und Tabakwaren – vornehmlich zu Zwecken der Preisstatistik – noch einmal differenzierter erfasst werden. Das Rotationsverfahren und die Beschränkung des Anteils der Haushalte, die Feinaufzeichnungen durchführen, dienen einerseits dazu, die Belastungen der Befragten durch die Aufzeichnungen zu verringern. Zum anderen ermöglicht die quartalsweise Aufzeichnung der Einnahmen und Ausgaben auch einen früheren ersten Zugang zu den anonymisierten Mikrodaten über Einnahmen und Ausgaben, als er bei ausschließlicher Verfügbarkeit der anonymisierten Gesamtstichprobe gegeben wäre.³⁾ Nachteil dieses Verfahrens ist natürlich, dass die Quartalsstichproben von Saisoneffekten beeinträchtigt sind.

2 Anonymisierte Mikrodaten

Das Statistische Bundesamt stellt der Wissenschaft Mikrodaten der EVS zur Verfügung, die faktisch anonymisiert sind. Gegenwärtig sind anonymisierte Mikrodatenfiles für die Erhebungen 1988, 1993, 1998 und 2003 verfügbar. Basis des anonymisierten Datensatzes ist eine zufällig gezogene Substichprobe von 80 – 98 Prozent der Originalstichprobe. Der Umfang der Substichprobe hängt von der Zahl der Erhebungsteile, die weitergegeben werden, ab.⁴⁾

Im Zuge der Anonymisierung werden vom Statistischen Bundesamt teilweise Zusammenfassungen bei ausgesuchten Merkmalen vorgenommen, so dass die univariate Verteilung nicht weniger als 5 000 Fälle (hochgerechnet auf die Gesamtbevölkerung) umfasst. Darüber hinaus werden bei ausgewählten Einkommens-, Vermögens- und Ausgabenbeträgen die fünf niedrigsten und die fünf höchsten Ausprägungen mit dem Mittelwert dieser Ausprägungen ausgewiesen sowie die übrigen Ausprägungen im untersten und obersten Dezil dieser Merkmale mit einem Zufallsfehler von bis zu +/- 1 % des jeweiligen Wertes überlagert.

2) So konnten die anonymisierten Mikrodaten des Einführungsinterviews für die Erhebung 2003 bereits im Frühjahr 2004 bereitgestellt werden, die anonymisierte Gesamtstichprobe war erst im September 2005 verfügbar. – 3) Das sog. Grundfile 3, das die anonymisierten Angaben aus dem Einführungsinterview und den ersten beiden Quartalen der Erhebung 2003 enthält, konnte bereits im Herbst 2004 bereitgestellt werden. – 4) Für detaillierte Informationen siehe <http://www.destatis.de/download/micro/mikrodaten.pdf> (Abrufdatum 20. 10. 2005)

3 Einkommensangaben in der EVS

Die einkommensbezogenen Daten der EVS stellen nicht den Schwerpunkt des vorliegenden Beitrags dar, gleichwohl sollen sie in einer kurzen Skizze beschrieben werden. Die EVS erheben grundsätzlich sehr detaillierte Daten zu allen Einkommensarten privater Haushalte:

Die Einkommen aus abhängiger Beschäftigung umfassen z. B. laufende Bezüge, Arbeitgeberzuschüsse zu freiwilliger oder privater Kranken- und Pflegeversicherung, Einmalzahlungen (Weihnachts-, Urlaubsgeld), Vermögenswirksame Leistungen, Abfindungen und Sachbezüge (in sehr tiefer Gliederung und Detaillierung).

Selbstständigeneinkommen werden aufgegliedert in Sachentnahmen, Einkommen aus land- und forstwirtschaftlicher Tätigkeit sowie Einkommen aus Gewerbebetrieb oder aus freiberuflicher Tätigkeit.

Der Bereich Transfereinkommen ist ebenfalls tief gegliedert: Grundsätzlich wird differenziert zwischen öffentlichen und nicht-öffentlichen Transferzahlungen. Unterschieden werden bei den öffentlichen Transferzahlungen z. B. Einkommen aus gesetzlichen Renten, Unfallrenten, Renten von berufsständischen Versorgungswerken und Zusatzversorgungskassen. Differenziert erfasst werden auch sonstige Transferzahlungen von Krankenkassen (Krankengeld, sonstige Übertragungen von Krankenkassen, Mutterschaftsgeld) sowie sonstige Leistungen der Sozialversicherung (Arbeitslosengeld, Pflegegeld, Altersteilzeitgeld) und steuerfinanzierte Transferzahlungen (z. B. Arbeitslosenhilfe, laufende Hilfe zum Lebensunterhalt, BaFÖG, Kindergeld, Unterhaltsvorschussleistungen, Erziehungsgeld, Wohngeld, Eigenheimzulage). Als nicht-öffentliche Transferzahlungen werden z. B. Werks- bzw. Betriebsrenten, Erstattungen und Leistungen privater Versicherungen, Kapitalauszahlungen aus Erbschaften und Auszahlungen privater Alters-, Pensions- und Sterbekassen klassifiziert.

Die Vermögenseinkommen differenzieren zwischen den Einnahmen aus Vermietung und Verpachtung sowie Zinsen, Dividenden und sonstigen Ausschüttungen. Als Besonderheit der EVS ist zu beachten, dass den Haushalten, die in Wohneigentum leben, eine kalkulatorische Miete für ihre selbst genutzte Immobilie als Einkommen zugerechnet wird. In vergleichenden Analysen mit anderen Datenquellen müssen diese unterstellten Eigentüermieten ggf. herausgerechnet werden.

Die Haushaltersparnis kann auf der Basis der EVS-Daten als Differenz zwischen der Summe der genannten Einkommen und den gesamten Konsumausgaben berechnet werden.

4 Vermögensangaben in der EVS

In der EVS werden Immobilienvermögen und Geldvermögen der privaten Haushalte erfasst. Gebrauchsvermögen und Betriebsvermögen werden dagegen nicht vollständig berücksichtigt. Hinsichtlich der Betriebsvermögen ist zu differenzieren zwischen Beteiligungen an Kapitalgesellschaften (Aktien, GmbH-Anteile, Genossenschaftsanteile), die unter „Aktien“ oder „Sonstige

Wertpapiere und Beteiligungen“ subsumiert und erhoben werden, und dem Besitz von bzw. Beteiligungen an Personengesellschaften, der nicht erhoben wird. Hinsichtlich der Gebrauchsvermögen berücksichtigt die EVS nur einen kleinen Teil langlebiger Konsumgüter, vorwiegend Kraftfahrzeuge und Unterhaltungselektronik. Hier wird aber lediglich das Vorhandensein dieser Vermögensgegenstände im Haushalt, nicht aber ihr Wert erfasst. Eine Auswertung bezüglich der durchschnittlichen Höhe und der Verteilung der Gebrauchsvermögen ist damit auf der Basis der EVS nicht möglich.

Hinsichtlich der Geldvermögen liefert die EVS prinzipiell nach Anlageformen gut differenzierte Daten, wenngleich auch hier wichtige Vermögenskategorien nicht einbezogen sind. Nicht berücksichtigt werden kurzfristige bzw. hoch liquide Anlageformen (Bargeld, Sichteinlagen)⁵⁾ sowie Guthaben, die im Rahmen betrieblicher kapitalgedeckter Altersvorsorgesysteme erworben wurden. Die übrigen Geldvermögensbestände werden dagegen in recht tiefer Gliederung erfasst. Unterschieden wurden in der Erhebung 2003 im Einzelnen Bausparverträge, Sparguthaben, sonstige Anlagen bei Banken und Versicherungen, Aktien, Aktienfonds, Rentenwerte, sonstige Wertpapiere und Beteiligungen sowie Lebensversicherungen, Ausbildungsversicherungen, Aussteuerversicherungen, Sterbegeldversicherungen und Ausleihungen von Privaten an Private. Auf der Schuldenseite werden im Bereich des Geldvermögens Konsumentenkredite erhoben.

Auch zu den Immobilienvermögen der privaten Haushalte enthält die EVS recht differenzierte Angaben. So enthalten die Daten Angaben darüber, ob es sich bei den Immobilien um unbebaute Grundstücke, Einfamilienhäuser, Zweifamilienhäuser, Häuser mit drei oder mehr Wohnungen oder sonstige Immobilien handelt. Für jede Immobilienart ist darüber hinaus die Information verfügbar, ob es sich um eine geerbte bzw. geschenkte oder um eine gekaufte bzw. gebaute Immobilie handelt. Leider fehlt in den Daten eine eindeutige Zuordnung zwischen dem angegebenen Immobilienbesitz und seiner Nutzungsart. Zwar sind Angaben dazu verfügbar, ob der Haushalt Eigentümer oder Mieter seiner Erstwohnung und – soweit vorhanden – seiner Zweitwohnung ist, diese Angaben sind aber nicht mit den sonstigen Angaben zum Immobilienbesitz verknüpft. Die Auswertungen bezüglich der Immobilienwerte sind ebenfalls begrenzt, da nur für alle Immobilien in der Summe ein geschätzter Verkehrswert ausgewiesen wird, aber keine Einzelwerte, wenn mehrere Immobilien vorhanden sind. Dazu korrespondierend werden auch die Restschulden nur in einer Summe ausgewiesen. Einheitswerte werden ebenfalls erhoben: In der Vergangenheit wurden die Einheitswerte bei fehlenden Angaben zu Verkehrswerten vom Statistischen Bundesamt auch zur Imputation von Verkehrswerten genutzt.

5 Probleme der Vermögensdaten in der EVS

Eine Analyse der Vermögensdaten in der EVS muss einige Besonderheiten der Stichprobe berücksichtigen, die vor allem die zeitliche Vergleichbarkeit der Erhebungen betreffen. Im Einzelnen sind folgende Problempunkte zu nennen:

5) Sichteinlagen werden zwar grundsätzlich erfasst, die Daten werden aber in den vom Statistischen Bundesamt herausgegebenen Mikrodatensätzen nicht weitergegeben.

- *Unterschiede* in den Abgrenzungen der Vermögenskategorien: Bedauerlicherweise weisen die Abgrenzungen einzelner Vermögenskategorien in verschiedenen Erhebungsjahren Unterschiede auf, sodass für einen Zeitvergleich erhebungsübergreifende und in geringerem Maße als in den Einzelerhebungen differenzierte Vermögenskategorien gebildet werden müssen. Dies betrifft im Wesentlichen den Bereich der Wertpapiere und Vermögensbeteiligungen, in dem z. B. in der Erhebung 2003 Aktien und Aktienfonds getrennt erhoben wurden, während letztere in den Vorjahren dem Aggregat „Sonstige Wertpapiere und Vermögensbeteiligungen“ zugeschlagen wurden. Umgekehrt wurden etwa 1993 Anteile an Immobilienfonds separat ausgewiesen, die dann in späteren Erhebungen nicht mehr getrennt erhoben wurden. Ein weiteres Beispiel für Abgrenzungsprobleme beim Zeitvergleich der EVS-Daten stellt eine in 2003 von den Vorerhebungen abweichende Definition der Restschuld aus Immobilienkrediten gegenüber den Vorerhebungen dar. Der Nutzer der Mikrodatensätze musste sich also bisher sehr detailliert und kritisch mit der Vergleichbarkeit der in den einzelnen Jahren erhobenen Vermögensaggregate auseinandersetzen. Für die Zukunft wäre hier etwas mehr Stetigkeit in den Abgrenzungen wünschenswert, wenngleich Änderungen im Anlageverhalten auch künftig durch Anpassungen in der Erhebungsstruktur Rechnung zu tragen sein wird.
- *Erhebungszeitpunkte*: Eine zeitvergleichende Betrachtung wird auch durch Unterschiede in den Erhebungszeitpunkten erschwert: Die Grundvermögen wurden in der EVS 1993 im Schlussinterview (zum 31. 12. 1993), 1998 und 2003 jeweils im Einführungsinterview (zum 1. 1. des Erhebungsjahres) erhoben. Die Geldvermögen wurden 1993 im Schlussinterview (zum 31. 12. 1993), 1998 im Rahmen der vierteljährlichen Anschreibungen (jeweils zum Ende eines Quartals) und 2003 im Einführungsinterview (zum 1. 1. 2003) erhoben. Daraus ergeben sich näherungsweise gleiche Zeiträume von durchschnittlich rund 4,5 Jahren zwischen den Erhebungszeiträumen 1993, 1998 und 2003 für das Geldvermögen, während in Bezug auf das Grundvermögen nur 4 Jahre zwischen der EVS 1993 und der EVS 1998, aber 5 Jahre zwischen der EVS 1998 und der EVS 2003 liegen.
- *Fehlende Werte*: Ein wesentliches Problem für den Zeitvergleich stellen die unterschiedliche Häufigkeit und die unterschiedliche Behandlung fehlender Vermögenswerte in den Mikrodatensätzen dar. Während die Stichprobe 1993 keine fehlenden Werte ausweist, enthält die Stichprobe 1998 fehlende Werte im Geldvermögen: Das Vorhandensein einer Vermögenskategorie wird in diesen Fällen bejaht, ohne dass ein Vermögenswert angegeben wird. Die Stichprobe 2003 weist fehlende Werte sowohl für das Geldvermögen als auch für das Immobilienvermögen aus, darüber hinaus enthält sie auch Fälle ohne jegliche Angaben zum Vermögen (vgl. zum Vorhandensein fehlender Werte im Geld- und Immobilienvermögen Tabelle 2 und 3). Die Unterschiede zwischen den Stichproben in Bezug auf fehlende Werte sind zum einen auf mögliche Unterschiede in der Erfassungsqualität zurückzuführen, da in früheren Erhebungen zu einem größeren Anteil persönliche Interviews statt schriftlicher Befragungen durchgeführt wurden. Zum anderen wurde von Seiten des Statistischen Bundesamtes in den letzten beiden Wellen zunehmend auf die Imputation fehlender Werte verzichtet. In der EVS 2003 wurden lediglich noch Imputationen von Rückkaufswerten bei Lebensversicherungen durchgeführt, sofern kein Rückkaufswert angegeben wurde, aber andere Einzelangaben zu den Versicherungsverträgen vorhanden waren. Der Vergleich der einzelnen Stichproben wird dadurch erschwert, dass die Imputation fehlender Werte in den früheren Erhebungen durch das Statistische Bundesamt nicht durchgehend dokumentiert ist.

- *Untererfassung*: Ein weiteres grundsätzliches Problem von Vermögenserhebungen durch Befragung privater Haushalte besteht in der tendenziellen Untererfassung der tatsächlich vorhandenen Vermögen. Die Ursachen dafür sind vielschichtig und reichen von der ungewollten Nichtberücksichtigung (z. B. Vergessen geringfügiger Sparguthaben oder der im Rahmen der so genannten Vermögenswirksamen Leistungen angesammelten Guthaben) über Bewertungsprobleme (z. B. beim Immobilienvermögen) bis zur bewussten Falschangabe, weil den Befragten die Vertraulichkeit der Daten nicht gesichert erscheint und z. B. steuerliche Konsequenzen korrekter Angaben befürchtet werden.

2. Angaben zu Geldvermögen in der EVS 2003							
Merkmal	Bausparverträge	Sparguthaben	Sonstige Anlagen bei Banken und Sparkassen	Wertpapiere	Verliehenes Geld an Privatpersonen	Konsumenkredite	Versicherungen
Anzahl Haushalte (ungewichtet)							
Nicht vorhanden	29 091	13 408	32 405	30 654	50 636	46 965	20 718
Vorhanden	27 384	42 665	23 208	25 845	4 899	9 565	35 640
Vorhanden, ohne weitere Angaben	155	432	306	165	27	89	317
Ohne Angaben	513	638	1 224	479	1 581	524	468
Summe	57 143	57 143	57 143	57 143	57 143	57 143	57 143
Anteil an allen Haushalten (ungewichtet)							
Vorhanden, ohne weitere Angaben	0,3 %	0,8 %	0,5 %	0,3 %	0,1 %	0,2 %	0,6 %
Ohne Angaben	0,9 %	1,1 %	2,1 %	0,8 %	2,8 %	0,9 %	0,8 %

Quelle: EVS, ZEW-Berechnungen.

3. Angaben zu Immobilienvermögen in der EVS 2003		
Merkmal	Anzahl Haushalte (ungewichtet)	Anteil an allen Haushalten (ungewichtet) %
Angaben zu Immobilienvermögen		
ohne Immobilien	23 205	40,61
mit Immobilien	32 809	57,42
– ohne Angabe eines Verkehrswertes	1 369	2,40
– ohne jegliche weitere Angaben	850	1,49
keine Angaben	279	0,49
Summe	57 143	100
Angaben zu Hypotheken		
ohne Hypotheken	19 351	33,86
mit Hypotheken	20 233	35,41
mit Hypotheken ohne weitere Angaben	195	0,34
keine Angaben	17 364	30,39
Summe	57 143	100

Quelle: EVS, ZEW-Berechnungen.

Allerdings fällt es nicht leicht, das Ausmaß der Untererfassung zu quantifizieren, da eine exakt vergleichbare Referenzgröße fehlt. Üblicherweise wird zum Vergleich die Finanzierungsrechnung der Deutschen Bundesbank herangezogen. Diese umfasst eine aggregierte Vermögensrechnung für den Sektor der privaten Haushalte inklusive der privaten Organisationen ohne Erwerbszweck. Sie ergänzt das Bild der auf die Erstellung von Gütern und Dienstleistungen und den Einkommenskreislauf ausgerichteten Volkswirtschaftlichen Gesamtrechnungen um die Darstellung derjenigen Vorgänge, die in der Finanzsphäre stattfinden. In ihrer sektoralen Gliederung folgt sie dem Europäischen System der Volkswirtschaftlichen Gesamtrechnungen (ESVG)

1995: Dementsprechend werden dem Sektor der privaten Haushalte in der Finanzierungsrechnung auch Einzelunternehmer und Selbstständige, deren unternehmerische Aktivitäten sich nicht von den Transaktionen der Privatpersonen trennen lassen, zugerechnet. Ebenfalls dem Sektor der privaten Haushalte wird in dieser Abgrenzung der Erwerb privaten Wohneigentums zugerechnet, der in früherer Abgrenzung der VGR im Unternehmenssektor erfasst wurde.⁶⁾

Die Finanzierungsrechnung beruht im Gegensatz zur EVS auf Angaben der Finanzintermediäre über die bei ihnen unterhaltenen Guthaben und aufgenommenen Schulden, sie ist in Bezug auf Vollständigkeit und Genauigkeit daher als zuverlässiger anzusehen als die Ergebnisse von Haushaltsbefragungen zu den Vermögensverhältnissen der privaten Haushalte. Auch wenn im Rahmen der Finanzierungsrechnung Schätzungen vorgenommen werden müssen, kann sie daher als Richtgröße für den tatsächlich vorhandenen Geldvermögensbestand angesehen werden.

Gegenüber der EVS als wichtigster Mikrodatenquelle zur Verteilung der Vermögen im Sektor der privaten Haushalte weist die Finanzierungsrechnung allerdings eine Reihe von Abgrenzungsunterschieden aus, so dass die Ergebnisse der beiden Datenquellen nur in bereinigter Form miteinander verglichen werden können. Diese Abgrenzungsunterschiede betreffen – neben der Einbeziehung der bereits erwähnten privaten Organisationen ohne Erwerbszweck – insbesondere die Einbeziehung von Bargeld und Sichteinlagen, die Berücksichtigung von Ansprüchen aus der betrieblichen Altersversorgung und aus berufsständischen Versorgungswerken sowie die Einbeziehung von Altersrückstellungen in der Krankenversicherung. Darüber hinaus ist die Finanzierungsrechnung im Gegensatz zur EVS nach dem Nettoprinzip organisiert, d. h. dass intrasektorale Verbindlichkeiten und Forderungen saldiert ausgewiesen werden. In der EVS werden dagegen auch Ansprüche gegen andere private Haushalte im Inland als Vermögensbestandteile ausgewiesen.

4. Erfassungsgrad des Geldvermögens in der EVS relativ zum Ausweis der Finanzierungsrechnung			
Art des Geldvermögen	1993	1998	2003
Bankeinlagen (inkl. Bausparguthaben)	55,6 %	58,9 %	79,7 %
Rentenwerte und Geldmarktpapiere	32,2 %	19,8 %	18,4 %
Aktien	24,7 %	26,7 %	57,7 %
Sonstige Beteiligungen und Investment-Zertifikate	35,7 %	31,2 %	38,9 %
Versicherungen	94,6 %	68,1 %	63,3 %
Bruttogeldvermögen insgesamt	53,5 %	46,9 %	56,8 %
Konsumentenkredite	22,2 %	27,6 %	28,2 %
Nettogeldvermögen	56,3 %	48,4 %	59,0 %

Quelle: Deutsche Bundesbank, EVS, Arbeitsgemeinschaft für betriebliche Altersvorsorge, GDV, ZEW-Berechnungen. Ergebnisse der Finanzierungsrechnung näherungsweise korrigiert um enthaltene Ansprüche aus betrieblicher Altersvorsorge und Altersrückstellungen der privaten Krankenversicherung, ohne Bargeld und Sichteinlagen, ohne private Organisationen ohne Erwerbszweck, ohne sonstige Forderungen und Verbindlichkeiten. EVS-Daten nach Imputation fehlender Werte.

Tabelle 4 stellt die Ergebnisse eines Vergleichs zwischen den Ergebnissen der Finanzierungsrechnung und den aggregierten Ergebnissen der EVS der Jahre 1993 bis 2003 dar. Es ist ersichtlich, dass die Erhebungsgrade zwischen den Anlageformen deutlich variieren. Tendenziell in recht hohem Maße erfasst sind Bankeinlagen (inkl. Bausparguthaben) und Versicherungsguthaben, deutlich untererfasst sind beispielsweise Rentenwerte und Geldmarktpapiere. Über die Ursachen dieser Abweichung kann trefflich spekuliert werden: Neben den oben genannten Erhebungsproblemen auf Haushaltsebene sind hier die Abgrenzungsunterschiede in den Datenquellen und auch die oben bereits angeführte Abschneidegrenze in der EVS, die sehr hohe Einkommen und dementsprechend auch sehr hohe Vermögen ausblendet, anzuführen.

6) Vgl. zu methodischen Fragen der Finanzierungsrechnung Deutsche Bundesbank (2005a), zu aktuellen Ergebnissen siehe Deutsche Bundesbank (2005b).

In methodischer Hinsicht stellt die mutmaßliche erhebliche Untererfassung in der EVS den Datennutzer vor erhebliche Probleme: Er kann sich für eine Korrektur der Untererfassung durch Hochrechnung beispielsweise am Maßstab der Finanzierungsrechnung entscheiden, muss dann aber bei Verteilungsanalysen annehmen, dass das Ausmaß der Untererfassung über die gesamte Verteilung zumindest annähernd gleich ist. Für die Richtigkeit dieser Annahme gibt es bislang aber keinen empirischen Beleg.

6 Vergleich mit dem SOEP

Während der Vergleich der EVS-Daten mit der Finanzierungsrechnung der Deutschen Bundesbank auf eine erhebliche Untererfassung hindeutet, kommen andere Haushaltsbefragungen zu ähnlichen Ergebnissen. Aufschlussreich ist hier insbesondere der Vergleich mit dem Sozio-ökonomischen Panel (SOEP). Das SOEP ist eine Panel-Befragung, die seit 1984 vom Deutschen Institut für Wirtschaftsforschung (DIW) durchgeführt wird. Es beinhaltet Personen-, Haushalts- und Familiendaten, wobei Schwerpunkte der Erhebung auf den Bereichen Bevölkerung und Demographie, Ausbildung und Qualifikation sowie Arbeitsmärkte und berufliche Dynamik liegen. Zum Vergleich mit der EVS eignet sich insbesondere die in der Welle 2002 des SOEP erhobene persönliche Vermögensbilanz, in der Angaben zur Höhe des Geld-, Immobilien-, Betriebs- und Sachvermögens der privaten Haushalte erfragt wurden. Insgesamt umfasst der Datensatz 2002 Befragungsergebnisse für 23 892 Personen in 12 692 Haushalten.

Das Geldvermögen ist im SOEP allerdings im Vergleich zur EVS nur grob kategorisiert; unterschieden werden lediglich zwei Blöcke: Geldanlagen (z. B. Sparguthaben, Spar- oder Pfandbriefe, Aktien oder Investmentanteile) und Lebensversicherungen/private Rentenversicherungen/Bausparverträge. Außerdem wurde die Höhe der Restverschuldung durch Kredite erhoben. Geldanlagen, Konsumentenkredite und Sachvermögen werden im Unterschied zur EVS erst ab einer Untergrenze von 2 500 Euro je Haushalt erfasst. Im Gegensatz zur EVS bezieht das SOEP auch Betriebsvermögen in einer breiteren Abgrenzung und Sachvermögen (sofern es eine Untergrenze von 2 500 Euro überschreitet) ein: Sachvermögen ist definiert als Gold, Schmuck, Münzen oder wertvolle Sammlungen (ohne Berücksichtigung von Kraftfahrzeugen). Darüber hinaus sind im SOEP (wie auch in den Erhebungen 1993 und 1998 der EVS) die Restschulden aus Hypothekenkrediten ohne Zinsen definiert, während in der EVS 2003 die Summe der noch ausstehenden Rückzahlungen angegeben werden muss.

Schließlich unterscheidet sich die Zusammensetzung der Stichprobe: So sind im SOEP auch Haushalte mit hohem Einkommen repräsentiert, die in der EVS wegen Überschreitens der Einkommensgrenze nicht enthalten sind. Zudem sind Ausländerhaushalte im SOEP tendenziell besser repräsentiert.⁷⁾ Ein Vergleich beider Stichproben muss diese Unterschiede, soweit wie möglich, berücksichtigen. In die nachstehend ausgewiesenen Vergleichsrechnung (siehe Tabelle 5) wurden deshalb nur Haushalte mit einem monatlichen Nettoeinkommen von weniger als 18 000 Euro (der Abschneidegrenze in der EVS) und mit deutschem Haushaltsvorstand einbezogen. Außerdem wurden die Vermögensabgrenzungen weitgehend vereinheitlicht und die Vermögensuntergrenzen des SOEP auch in der EVS verwendet.

⁷⁾ Vgl. zu einer ausführlichen Diskussion der Strukturunterschiede in den Stichproben Becker et al. (2003), S. 63 f.

5. Mittelwerte der Vermögen SOEP und EVS im Vergleich (1 000 Euro)								
Merkmal	SOEP			EVS			Relation EVS/SOEP	
	Alle	Besitzquote	Besitzer	Alle	Besitzquote	Besitzer	Alle	Besitzer
Bruttoverm.	157,6	75 %	210,9	161,8	86 %	187,1	103 %	89 %
Schulden	24,2	29 %	82,9	27,9	36 %	78,4	115 %	95 %
Nettoverm.	133,4	71 %	188,7	133,9	83 %	161,8	100 %	86 %
Immob. (brutto)	112,8	41 %	276,9	120,9	49 %	245,4	107 %	89 %
Hypotheken	20,3	19 %	105,7	26,6	28 %	95,8	131 %	91 %
Immob. (netto)	92,5	40 %	230,9	94,3	48 %	197,8	102 %	86 %
Bruttogeldv.	31,6	67 %	46,9	40,9	84 %	48,8	129 %	104 %
Konsum.kredite	3,9	14 %	28,8	1,3	11 %	11,7	34 %	41 %
Nettogeldv.	27,7	63 %	47,8	39,6	80 %	50,1	143 %	105 %
Geldanlagen	16,9	47 %	36,2	26,4	65 %	40,4	157 %	112 %
Versicherung.	14,8	51 %	29,2	14,5	67 %	21,6	98 %	74 %
Betriebsverm.	10,9	6 %	194,0					
Sachverm.	2,4	11 %	22,5					

Quelle: SOEP, ZEW-Berechnungen. Gleiche Abgrenzung für Untergrenzen und Einkommensintervall, nur Haushalte mit deutlichem Haushaltsvorstand, Besitzquoten bei Nettowerten ohne negative Werte. Nach Imputation fehlender Werte.

Im Ergebnis zeigen sich im Hinblick auf die mittlere Höhe des Nettovermögens über alle Haushalte nahezu identische Werte in beiden Stichproben. Da in der EVS ein etwas höherer Schuldenstand ausgewiesen wird, liegt das durchschnittliche Bruttovermögen der EVS in dieser Abgrenzung um rund 3 % über dem Bruttovermögen der Haushalte im SOEP. Zu berücksichtigen ist hier allerdings die umfassendere Definition des Gesamtvermögens im SOEP, die auch Betriebsvermögen in weiterer Abgrenzung und – quantitativ allerdings nicht stark ins Gewicht fallende – Sachvermögen umfasst. In einzelnen Vermögenskategorien und bei Betrachtung der Besitzquoten zeigen sich gleichwohl größere Unterschiede. Da das SOEP – mit Ausnahme der Konsumentenkredite – in allen Vermögenskategorien geringere Besitzquoten als die EVS ausweist, sind im Allgemeinen die Mittelwerte des Vermögens derjenigen Haushalte, die die entsprechenden Vermögenskategorien wirklich besitzen, größer als in der EVS. Besonders groß sind die Unterschiede bei der Erfassung der Geldanlagen (im SOEP definiert als Sparguthaben, Spar- oder Pfandbriefe, Aktien oder Investmentanteile) und bei Versicherungen (inkl. Bausparverträge). Für diese Aggregate – die für den Vergleich mit der EVS in diesem Datensatz entsprechend definiert wurden, zeigen sich mit 47 % bzw. 51 % im SOEP deutlich geringere Besitzquoten als in der EVS (65 % für Geldanlagen bzw. 67 % für Versicherungen).⁸⁾

Die Geldvermögen werden daher bei Betrachtung über alle Haushalte in der EVS höher abgebildet als dies im SOEP der Fall ist. Die Höhe der Geldvermögen je Besitzer stimmt dagegen – in der hier vorgenommenen Aufbereitung des Datensatzes – weitgehend mit dem Niveau der Geldvermögen in der EVS überein. Allerdings ist zu berücksichtigen, dass in der EVS im Geldvermögen grundsätzlich auch Unternehmensbeteiligungen in Form von Anteilen an Personengesellschaften, Genossenschaften und Gesellschaften mit beschränkter Haftung enthalten sind, die im SOEP als Betriebsvermögen ausgewiesen werden dürften.

Das Immobilienvermögen ist im SOEP mit einer Eigentumsquote von 41 % im Vergleich zur EVS (49 %) tendenziell untererfasst. Vermutlich auch wegen der unterschiedlichen Definition der Restschulden ergeben sich aber bei Betrachtung über alle Haushalte vergleichbare Nettowerte. Betrachtet man die Vermögensverteilung, so ist – vor allem wohl zurückgehend auf die geringere Erfassung des Bruttogeldvermögens und die höheren Konsumentenschulden – ein

8) Beim Vergleich dieser Besitzquoten ist die unterschiedliche Abgrenzung der Aggregate – das Aggregat der Versicherungen im SOEP enthält auch Bausparverträge – zu berücksichtigen.

höheres Maß an Ungleichverteilung im SOEP zu erkennen. Zusammenfassend kann der Grad an Übereinstimmung zwischen den Resultaten beider Erhebungen – angesichts der großen Probleme, mit denen Vermögenserhebungen üblicherweise konfrontiert sind – jedoch als recht hoch eingeschätzt werden.

7 Ergebnisse aktueller Analysen zur Vermögensverteilung auf der Basis der EVS

Im Rahmen des 2. Armuts- und Reichtumsberichts der Bundesregierung hat das ZEW auf der Basis hauptsächlich der EVS-Daten die Vermögen der privaten Haushalte und ihre Verteilung analysiert.⁹⁾ Dabei konnten folgende Ergebnisse erzielt werden:

In der Erhebung 2003 weisen die privaten Haushalte im Durchschnitt für das gesamte Bundesgebiet ein materielles Nettogesamtvermögen (definiert als Summe von Immobilienvermögen und Geldvermögen) von rund 133 000 Euro aus (siehe Tabelle 6). Im Verlauf der letzten zehn Jahre sind die Nettovermögen der Haushalte damit im Durchschnitt in nominaler Rechnung um knapp 26 % angestiegen. Der Median des Nettovermögens liegt mit 50 000 Euro allerdings erheblich niedriger. Zudem zeigen sich erhebliche Unterschiede zwischen West- und Ostdeutschland. Während die westdeutschen privaten Haushalte im Durchschnitt über ein Geld- und Immobilienvermögen von rund 149 000 Euro verfügen, umfassen die Vermögen der ostdeutschen Haushalte mit knapp 60 000 Euro im Durchschnitt nur 40 % des Vermögens der westdeutschen Haushalte. Jedoch hat sich der Abstand zwischen ost- und westdeutschen Haushalten im Zeitverlauf erheblich verringert. Die Vermögen der ostdeutschen Haushalte sind in den vergangenen 10 Jahren um rund 63 % gewachsen. Mit knapp 19 % blieben die Zuwachsraten der westdeutschen Privatvermögen deutlich dahinter zurück.

6. Mittelwert und Median der Gesamtvermögen je Haushalt (Geld- und Immobilienvermögen) im Vergleich (in 1 000 Euro)						
Merkmal	1993		1998		2003	
	Mittelwert	Median	Mittelwert	Median	Mittelwert	Median
Deutschland						
Bruttovermögen	121,1	35,6	133,7	45,6	161,3	67,0
Schulden	15,0	0	20,0	0	27,9	0
Nettovermögen	106,2	32,4	113,7	38,5	133,4	49,8
Alte Bundesländer						
Bruttovermögen	143,3	79,4	151,3	74,8	179,0	93,5
Schulden	17,9	0	22,2	0	30,2	0
Nettovermögen	125,4	60	129,2	56,1	148,8	63,6
Neue Bundesländer						
Bruttovermögen	40,1	10,6	56,3	16,8	76,1	25,2
Schulden	3,7	0	10,6	0	16,6	0
Nettovermögen	36,4	10,1	45,6	15,4	59,6	21,8

Quelle: EVS, ZEW-Berechnungen. Ohne Berücksichtigung privat verliehener Gelder, Bargeld und Sichteinlagen. Nach Imputation fehlender Werte.

9) Siehe dazu ausführlicher Ammermüller, Weber, Westerheide (2005).

Eine Analyse der Vermögensverteilung für Deutschland zeigt eine erhebliche Ungleichverteilung des privaten Vermögens. Die Haushalte in der unteren Hälfte der Vermögensverteilung verfügen über etwas weniger als 4 % des gesamten Nettovermögens, während die 20 % vermögensstärksten Haushalte rund zwei Drittel des gesamten privaten Geld- und Immobilienvermögens auf sich vereinen. Bei getrennter Betrachtung der Verteilung in Ost- und Westdeutschland wird erkennbar, dass die Vermögensverteilung in Ostdeutschland etwas ungleichmäßiger als in Westdeutschland ist. Allerdings haben sich die Unterschiede zwischen der westdeutschen und der ostdeutschen Vermögensverteilung im Zeitverlauf deutlich verringert. Während die ostdeutsche Vermögensverteilung tendenziell gleichmäßiger geworden ist, ist für die westdeutsche Verteilung eine Tendenz zu einer stärkeren Ungleichverteilung festzustellen.

So haben die Vermögen der Haushalte in Westdeutschland in der Zehnjahresperspektive im unteren Drittel der Verteilung – auf allerdings niedrigem Niveau – in prozentualer Betrachtung deutlich abgenommen. In der Mitte der Verteilung sind im Allgemeinen geringe Zuwächse zu verzeichnen, während bei den höheren Vermögen stärkere Zuwächse zu konstatieren sind. In Ostdeutschland ist dagegen in der Zehnjahresperspektive ein substanzieller Aufbau von Vermögen über einen breiten Bereich der Vermögensverteilung zu beobachten. Die Zunahme am oberen Rand der Verteilung fällt dagegen relativ gering aus.

Die Entwicklung der gesamtdeutschen Verteilung weicht deutlich von der Entwicklung der westdeutschen Verteilung ab, was angesichts des hohen Anteils der westdeutschen Vermögen am gesamtdeutschen privaten Vermögen auf den ersten Blick überrascht. Ursächlich für die starken Abweichungen zwischen west- und gesamtdeutscher Verteilung sind die erheblichen Zuwächse des Vermögens in der oberen Hälfte der ostdeutschen Vermögensverteilung, die – wegen des geringeren Durchschnittsniveaus der ostdeutschen Vermögen – insbesondere die Mitte der gesamtdeutschen Vermögensverteilung beeinflussen.

Eine nach sozialen Gruppen differenzierende Analyse der Vermögensverteilung zeigt erhebliche Unterschiede in den Durchschnittsvermögen: Arbeitnehmerhaushalte verfügen im bundesdeutschen Durchschnitt über ein Vermögen von rund 120 000 Euro, Selbstständige dagegen über rund 300 000 Euro privates Geld- und Immobilienvermögen. Die durchschnittlichen Vermögen der Arbeitnehmerhaushalte haben im Zeitraum von 1993 bis 2003 mit durchschnittlich rund 20 % rund doppelt so stark zugenommen wie die Privatvermögen der Selbstständigenhaushalte. Der stärkste durchschnittliche Vermögenszuwachs ist bei den Nichterwerbstätigenhaushalten festzustellen. Hinter dieser Entwicklung verbergen sich vor allem deutlich steigende Vermögen von Rentnern und Pensionären. Die Vermögen von Arbeitslosen stagnieren dagegen in der langfristigen Betrachtung. Vom gesamten privaten Geld- und Immobilienvermögen entfallen rund 46 % auf Arbeitnehmerhaushalte, 16 % auf Selbstständigenhaushalte und 38 % auf Nichterwerbstätigenhaushalte. Unter den Nichterwerbstätigen stellen Rentner und Pensionäre mit rund 35 % Anteil am Privatvermögen die größte Gruppe dar.

Literatur

Ammermüller, Andreas, Andrea Maria Weber und Peter Westerheide (2005): Die Entwicklung und Verteilung des Vermögens privater Haushalte unter besonderer Berücksichtigung des Produktivvermögens, Bundesministerium für Gesundheit und Soziale Sicherung, Mannheim. Download unter <http://www.zew.de>.

Becker, I. et al. (2003): A comparison of the main household income surveys for Germany: EVS and SOEP, in: Irene Becker/Richard Hauser (Hrsg.): Reporting on income distribution and poverty : perspectives from a German and a European point of view, Berlin.

Deutsche Bundesbank (2005a): Ergebnisse der gesamtwirtschaftlichen Finanzierungsrechnung für Deutschland 1991 bis 2004. Statistische Sonderveröffentlichung 4. Download unter http://www.bundesbank.de/stat/download/stat_sonder/statso4.pdf (Abrufdatum 20. 10. 2005).

Deutsche Bundesbank (2005b): Vermögensbildung und Finanzierung im Jahr 2004. In Monatsbericht Juni 2005, S. 15ff. Download unter http://www.bundesbank.de/download/volkswirtschaft/mba/2005/200506mba_Vermoegensbildung_2004.pdf (Abrufdatum 20. 10. 2005).

Autoren- und Kontaktangaben:

Dr. Peter Westerheide, Zentrum für Europäische Wirtschaftsforschung GmbH (ZEW), Forschungsbereich Internationale Finanzmärkte und Finanzmanagement, Postfach 10 34 43, 68034 Mannheim, E-Mail: westerheide@zew.de

Krankenhaus-Diagnosestatistik – Scientific-Use-File und beispielhafte Analysen*)

Gabriele Philipp, Olaf Schoffer

Basierend auf der im Bundesstatistikgesetz geregelten „faktischen Anonymität“ werden im Forschungsdatenzentrum der Statistischen Ämter der Länder Mikrodaten der amtlichen Statistik zur wissenschaftlichen Nutzung aufbereitet. Die Aufbereitung von Daten der Gesundheitsstatistik erfolgt dabei als ein Schwerpunkt am Regionalen Standort Kamenz. Insbesondere schließt dies die Erstellung eines Scientific-Use-Files der Krankenhaus-Diagnosestatistik ein. Nachstehend werden daher die Krankenhausstatistik und das daraus abgeleitete Scientific-Use-File der Diagnosestatistik ausführlich vorgestellt. Ergänzt wird diese Darstellung um einige beispielhafte Analysen, um das dieser Datenbasis innewohnende Analysepotenzial zu demonstrieren.

1 Einleitung

Gemäß dem Bundesstatistikgesetz (BStatG) „dürfen vom Statistischen Bundesamt und den statistischen Ämtern der Länder Einzelangaben an Hochschulen und sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung übermittelt werden, wenn die Einzelangaben nur mit unverhältnismäßig großem Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können“. ¹⁾ Auf Grundlage dieser Regelung zur „faktischen Anonymität“ ermöglichen die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder der Wissenschaft einen privilegierten Zugang zu Mikrodaten der amtlichen Statistik. ²⁾

In den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder werden verschiedene Statistiken fachlich zentralisiert aufbereitet und der Wissenschaft als Einzeldaten zur Verfügung gestellt. ³⁾ Im Forschungsdatenzentrum der Statistischen Ämter der Länder (FDZ/L) stehen dabei insbesondere regional tief gegliederte Einzeldaten für wissenschaftliche Analysen zur Verfügung.

Der regionale Standort Kamenz des FDZ/L bearbeitet als fachlichen Schwerpunkt die Statistiken des Gesundheits- und Sozialwesens, die jedoch an jedem anderen regionalen Standort zur Verfügung gestellt werden können. Hier werden die Einzeldaten aller Bundesländer aus den Bereichen Krankenhaus- und Todesursachen-Statistik sowie Sozialhilfe- und Jugendhilfestatistik gesammelt und aufbereitet. Dies schließt nicht zuletzt die Erstellung standardisierter faktisch anonymisierter Mikrodatenfiles (Scientific-Use-Files) ein.

*) Dieser Aufsatz erschien in leicht veränderter Fassung zuerst in DRV-Schriften, Band 55/2005 „Bericht vom zweiten Workshop des FDZ-RV vom 27. bis 29. Juni 2005 in Würzburg“. – 1) § 16 Abs. 6 BStatG. – 2) Vgl. Scheffler (2005). – 3) Für eine ausführliche Darstellung zu den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder vgl. Zühlke, Zwick, Scharnhorst, Wende (2004). Die gemeinsame Internet-Präsenz der Forschungsdatenzentren unter www.forschungsdatenzentrum.de enthält darüber hinaus Informationen zum gemeinsamen Datenangebot sowie über aktuelle Projekte und Veranstaltungen.

Ein wichtiges Vorhaben des Standortes Kamenz ist daher die Konzeption eines Scientific-Use-Files der Krankenhaus-Diagnosestatistik. Die Arbeiten dafür erfolgen in Zusammenarbeit mit dem Statistischen Bundesamt im Rahmen eines Anonymisierungsprojektes. Zusätzlich werden für die Krankenhaus-Diagnosestatistik in Form eines Forschungsprojektes am Standort Kamenz beispielhaft deskriptive Analysen durchgeführt, um das Potenzial dieser Daten verdeutlichen zu können. Die Ergebnisse beider Projekte werden nachfolgend vorgestellt.

Weitere aktuelle Projekte am Standort Kamenz sind die Anonymisierung der Jugendhilfestatistik für die Arbeit am Gastwissenschaftler-Arbeitsplatz sowie die Aufbereitung der Krankenhausstatistik (Grund- und Kostendaten) für die Analyse mittels kontrollierter Datenfernverarbeitung. In Vorbereitung ist außerdem die Aufbereitung der Einzeldaten für die Todesursachenstatistik aller Bundesländer, welche neben der Diagnosestatistik die zweite Schwerpunktstatistik für den Standort Kamenz bildet.

2 Diagnosestatistik als Teil der Krankenhausstatistik

Da die betrachtete Diagnosestatistik einen Teil der Krankenhausstatistik darstellt, folgt zunächst ein Überblick über deren allgemeine Eigenschaften. Die Krankenhausstatistik ist eine Totalerhebung der Krankenhäuser sowie der Vorsorge- oder Rehabilitationseinrichtungen. Sie enthält Informationen über organisatorische Einheiten, personelle und sachliche Ausstattung sowie erbrachte Leistungen in den jeweiligen Einrichtungen. Somit sind beispielsweise Aussagen über die Struktur der stationären Versorgung, die Zusammenhänge zwischen Morbiditäts- und Kostenentwicklung im Krankenhausbereich oder die regionale Häufung von Krankheitsarten möglich.

Damit bildet die Krankenhausstatistik die Planungsgrundlage vieler gesundheitspolitischer Entscheidungen. Beispielsweise dient sie zur Bedarfsermittlung an Krankenhausleistungen und technischer Ausstattung oder zur Untersuchung von Lücken in der medizinischen Versorgung. Ebenso ist sie Grundlage verschiedenster wissenschaftlicher Analysen, wie etwa Untersuchungen von Patientenströmen oder zur Entwicklung bestimmter Erkrankungen.

Die Krankenhausstatistik gliedert sich in drei Teile. Neben der Diagnosestatistik (Teil II) ermöglichen die Grunddaten (Teil I) mit Informationen zu technischer/personeller Ausstattung und erbrachten Leistungen der Einrichtungen sowie der Kostennachweis (Teil III) mit Informationen zu Sach-, Personal- und Ausbildungskosten ein umfassendes Bild der deutschen Krankenhauslandschaft.

Die Diagnosestatistik umfasst die Hauptdiagnosen sowie allgemeine Angaben zur Person entlassener oder verstorbener vollstationärer Patienten⁴⁾ aus Krankenhäusern. Als Hauptdiagnose wird die Diagnose angegeben, welche hauptsächlich die Dauer der stationären Behandlung beeinflusst bzw. den größten Teil medizinischer Leistungen verursacht hat. Sie wird seit dem Berichtsjahr 2000 nach der 10. Revision der Internationalen Klassifikation der Krankheiten (ICD-10)⁵⁾ verschlüsselt, zuvor kam die ICD-9 zur Anwendung.

4) Aus Gründen der Lesbarkeit wurde im Text zumeist die männliche Form gewählt, auch wenn sich die Angaben auf Angehörige beider Geschlechter beziehen. – 5) DIMDI (2000).

Die Diagnosestatistik wird zusammen mit den anderen Teilen der Krankenhausstatistik jährlich zum 31.12. erhoben. Mehrfach im Jahr vollstationär behandelte Patienten werden entsprechend mehrfach erfasst, aus jedem Krankenhausaufenthalt wird ein neuer Eintrag zusammengestellt. Die Diagnosestatistik ist somit fallbezogen – ein Rückschluss auf die Anzahl der Krankenhausaufenthalte eines bestimmten Patienten ist aufgrund fehlender eindeutiger personen-identifizierender Merkmale nicht möglich.

Die Daten der Krankenhauspatienten werden seit 1993 erfasst. Seit 2003 wird die Diagnosestatistik zusätzlich auch in Vorsorge- sowie Rehabilitationseinrichtungen mit 100 und mehr Betten erhoben.

Die beschriebenen Informationen über die behandelten Patienten werden um Angaben aus den Grunddaten der Krankenhausstatistik ergänzt, um leichter untersuchen zu können, in Einrichtungen welcher Art die Behandlungen erfolgten. Damit sind im Wesentlichen folgende Merkmale in der Diagnosestatistik enthalten:

Angaben zu den behandelten Patienten

- (Haupt-)Diagnose
- Wohnort, Geschlecht, Alter
- Verweildauer (mit Aufnahme- und Entlassungsdatum)
- Sterbefall (ja/nein)
- Operation im Zusammenhang mit Hauptdiagnose (ja/nein)
- Fachabteilung mit längster Verweildauer

Angaben zu den behandelnden Einrichtungen

- Standort der Einrichtung
- Art der Einrichtung, des Trägers und der Förderung
- Anzahl der Betten und der Fachabteilungen
- Anteil der Belegbetten

Die Erarbeitung sowie Komplettierung von Metadaten für die Diagnosestatistik findet zurzeit statt. Der Qualitätsbericht zu dieser Statistik liegt bereits vor.⁶⁾

3 Scientific-Use-File der Diagnosestatistik

In der ersten Nutzerumfrage zum Datenbedarf der Wissenschaft⁷⁾ zählt die Statistik der Krankenhäuser zu den 15 am häufigsten genannten Statistiken, welche weitgehend deckungsgleich zu den im Gründungsausschuss des Rates für Sozial- und Wirtschaftsdaten für den Aufbau des Forschungsdatenzentrums der Statistischen Landesämter als prioritär Genannten sind.

Daher wurde festgelegt, für die Diagnosestatistik als einem zentralen Bestandteil der Krankenhausstatistik ein Scientific-Use-File zu erstellen. Ein solches standardisiertes Mikrodatenfile ermöglicht es, auf Datenbedarf für wissenschaftliche Analysen schnell reagieren zu können. Vor-

6) http://www.destatis.de/download/qualitaetsberichte/qualitaetsbericht_diagnosekhp.pdf – 7) Vgl. Zühlke, Hetke (2003).

aussetzung für die Akzeptanz dieses Datenangebotes durch die Wissenschaft ist jedoch, dass das Analysepotenzial bei der dabei erforderlichen faktischen Anonymisierung der Einzeldaten bestmöglich erhalten bleibt.

Um dieses Ziel zu erreichen, erfolgt die Konzeption eines faktisch anonymisierten Mikrodatenfiles der Diagnosestatistik unter Einbeziehung eines wissenschaftlichen Beraterkreises. Auf Basis der Vorschläge dieses Beraterkreises und von Überlegungen sowie Analysen zur faktischen Anonymisierbarkeit ist das nachfolgend beschriebene Konzept für das Scientific-Use-File der Diagnosedaten von Krankenhauspatienten und -patientinnen entstanden.

Zentrale Elemente des Anonymisierungskonzeptes sind:

- der Erhalt der Vollerhebung,
- die Trennung der Angaben zu Patienten und zu Krankenhäusern in Form zweier Datensätze,
- die Beschränkung auf wenige „wichtige“ Variablen und
- die Vergrößerung der Merkmalswerte einiger Variablen.

Der Verzicht auf Ziehung einer Stichprobe zugunsten des Erhalts aller Fälle von im Berichtsjahr entlassenen Patienten ist ein wichtiger Schritt zur Erhaltung der Repräsentativität für die deutsche Krankenhauslandschaft. Im Unterschied zum Einzeldatenmaterial mussten aus Gründen der Geheimhaltung dennoch einige Fälle im Scientific-Use-File entfallen. Insbesondere sind das Fälle, deren Merkmale Hauptdiagnose, Verweildauer, Alter oder Geschlecht als Merkmalswerte „unbekannt“ aufweisen sowie Fälle, welche eine eindeutige Zuordnung zu bestimmten Personen erlauben. Der Anteil dieser Fälle, welche in Analysen ohnehin oft ausgeschlossen würden, liegt im Promillebereich. Dies führt dazu, dass die mit dem Scientific-Use-File berechneten Ergebnisse von den Publikationen der statistischen Ämter abweichen. Anhand der dokumentierten Anzahl entfallener Fälle ist es jedoch möglich, die Größenordnung der dadurch verursachten Änderung abzuschätzen.

Bei der Anonymisierung ist zunächst zu bedenken, welche Informationen besonders schutzwürdig sind. Für Krankenhäuser ist insbesondere zu vermeiden, dass Angaben zu ihrer wirtschaftlichen Situation bekannt werden. Für die Patienten hingegen ist hauptsächlich die Hauptdiagnose als Grund des jeweiligen Krankenhausaufenthaltes zu schützen, da diese durch Dritte, etwa Arbeitgeber oder Nachbarn, zum Nachteil des Patienten verwendet werden kann.

Um dieser Anforderung bei möglichst geringem Verlust an Analysepotenzial gerecht zu werden, erfolgt die Aufteilung der Angaben zu Patienten und zu Krankenhäusern auf zwei getrennte Datensätze. Somit wird die Identifikation sowohl von Patienten als auch von Krankenhäusern über zusätzliche Informationen erschwert. Der erste Datensatz beinhaltet hauptsächlich patientenbezogene Daten und soll vornehmlich epidemiologischen Untersuchungen dienen. Der zweite Datensatz fokussiert dagegen krankenhausbazogene Daten und orientiert sich damit stärker an gesundheitsökonomischen und -politischen Fragestellungen.

Als weitere Maßnahme zur Anonymisierung werden, gemäß vom wissenschaftlichen Beraterkreis vorgeschlagener Prioritäten, nur die wichtigsten Merkmale in das Scientific-Use-File übernommen. Für wissenschaftliche Analysen sind die Anzahl der Patienten sowie deren Verweildauer nach Hauptdiagnose (vierstelliger ICD-Schlüssel) von zentraler Bedeutung. Daher sind

1. Merkmale und Merkmalsausprägungen im Scientific-Use-File		
Merkmal	Merkmalsausprägungen	
	Patientenbezogener Datensatz	Krankenhausbezogener Datensatz
Hauptdiagnose	Vierstelliger ICD-10 Schlüssel	Vierstelliger ICD-10 Schlüssel
Verweildauer	0 Tage 1 Tag usw. in 1-Tages-Schritten bis 42 Tage 43 bis 49 Tage 50 bis 56 Tage 57 bis 63 Tage 64 bis 70 Tage 71 bis 84 Tage 85 bis 98 Tage 99 bis 140 Tage 141 bis 182 Tage mehr als 182 Tage	0 Tage 1 Tag usw. in 1-Tages-Schritten bis 42 Tage 43 bis 46 Tage 47 bis 53 Tage 54 bis 60 Tage 61 bis 67 Tage 68 bis 77 Tage 78 bis 91 Tage 92 bis 119 Tage 120 bis 161 Tage mehr als 161 Tage
Alter	0 bis unter 5 Jahre 5 bis unter 10 Jahre usw. in 5-Jahres-Schritten bis unter 95 Jahre 95 Jahre und älter	0 bis unter 3 Jahre 3 bis unter 8 Jahre usw. in 5-Jahres-Schritten bis unter 93 Jahre 93 Jahre und älter
Geschlecht	Männlich Weiblich	*
Patientenwohnort	AGS bis Kreisebene (5 Stellen) Ausland unbekannt	*
Behandlungsort	*	Nord (SH, HH, HB, NI, NW) Süd (BW, BY, HE, RP, SL) Ost (BE, BB, MV, SN, ST, TH)
Fachabteilung mit der längsten Verweildauer	*	Chirurgie Frauenheilkunde und Geburtshilfe Hals-Nasen-Ohren-Heilkunde Innere Medizin Kinderheilkunde Orthopädie Urologie Psychiatrie Sonstige unbekannt
Trägerschaft	*	öffentlich freigemeinnützig privat
Bettengrößenklasse	*	unter 100 Betten** 100 bis unter 200 Betten** 200 bis unter 500 Betten*** 500 und mehr Betten***

*) Merkmal im Datensatz nicht enthalten – **) für öffentliche Krankenhäuser in Ost werden diese zwei Klassen zu „unter 200 Betten“ zusammengefasst – ***) für private Krankenhäuser werden diese zwei Klassen zu „200 und mehr Betten“ zusammengefasst – – – Quelle: Forschungsdatenzentren der statistischen Ämter des Bundes und der Länder – Arbeiten am Anonymisierungskonzept.

die beiden Zielvariablen Verweildauer und Hauptdiagnose in beiden Datensätzen enthalten. Im patientenbezogenen Datensatz (DS I) werden sie ergänzt um Alter, Geschlecht und Wohnort des Patienten. Im krankenhausbezogenen Datensatz (DS II) werden die Zielgrößen erweitert um Alter des Patienten, Behandlungsgebiet, Fachabteilung mit der längsten Verweildauer sowie Trägerschaft und Bettengrößenklasse des Krankenhauses.

Neben den bereits beschriebenen Anonymisierungsmaßnahmen erfordern einige Merkmale eine Vergrößerung, also eine Gruppierung der Merkmalsausprägungen. So wird der Patientenwohnort nur noch bis auf Kreisebene angegeben, der ursprünglich achtstellige Amtliche Gemeindegemeinschaftsschlüssel (AGS) also auf fünf Stellen eingeschränkt. Für den Behandlungsort wird hingegen die Angabe des Bundeslandes aus dem AGS verwendet, um drei Gebiete (Nord, Süd und Ost) zu bilden. Die Merkmalsausprägungen von Fachabteilung mit der längsten Verweildauer und Bettengrößenklasse werden in Gruppen zusammengefasst. Die Merkmalsausprägungen von Verweildauer und Alter werden ebenfalls gruppiert, jedoch ist für diese die Gruppierung zwischen patienten- und krankenhausbezogenem Datensatz verschoben.

Der aus den beschriebenen Anonymisierungsmaßnahmen resultierende Aufbau des Scientific-Use-File mit Details bezüglich der Merkmalsausprägungen ist Tabelle 1 zu entnehmen.

Merkmale, welche durch die beschriebene Anonymisierung im Scientific-Use-File nicht oder nur vergrößert berücksichtigt werden, können über andere Zugangswege des FDZ/L genutzt werden. Dies sind insbesondere die On-Site-Nutzung an einem Gastwissenschaftler-Arbeitsplatz in den Räumen eines Statistischen Landesamtes sowie die kontrollierte Datenfernverarbeitung.

4 Beispielhafte Analysen der Diagnosestatistik

Das Statistische Landesamt des Freistaates Sachsen hat im Februar 2005 für das Forschungsprojekt „Beispielhafte Analysen der Krankenhaus-Diagnosestatistik“ einen Antrag zur Nutzung der Einzeldaten beim FDZ/L gestellt. Ziel des Projektes ist es, einen Eindruck davon zu vermitteln, welche Analysemöglichkeiten die Daten bieten sowie Anregungen für Analyseansätze zu geben, ohne dabei zu sehr in die Tiefe zu gehen. Anhand des zuerst vollständig vorliegenden Einzelmateriale des Jahres 2001 wurden beispielhafte Analysen, Grafiken und Fragestellungen erarbeitet, die zur Öffentlichkeitsarbeit, insbesondere auf Nutzerkonferenzen und Themenworkshops des FDZ/L verwendet werden.

Einige der im Rahmen dieses Projektes erarbeiteten Ergebnisse werden nachfolgend vorgestellt.⁸⁾ In der ersten Analyse wird die Wanderung von Krankenhaus-Patienten betrachtet. In Tabelle 2 sind Prozentanteile der Patienten angegeben, welche sich in dem Bundesland behandeln lassen, in dem sie wohnhaft sind, sowie die Prozentanteile derer, die zur Behandlung ihr Bundesland verlassen haben.

8) Die Analysen wurden anhand der Mikrodaten der Diagnosestatistik durchgeführt. Daher sei darauf hingewiesen, dass nicht alle der angesprochenen Fragestellungen mit dem Scientific-Use-File in der hier verwendeten Detailltiefe zu analysieren sind. Dafür stehen jedoch die erwähnten anderen Zugangswege des FDZ/L offen.

Auf eine ausführliche Interpretation der Ergebnisse soll an dieser Stelle verzichtet werden. Besonders häufig gibt es jedoch Patientenströme zwischen den Stadtstaaten und den umgebenden Bundesländern sowie aus Rheinland-Pfalz in die dort umgebenden Bundesländer.

Analog zu der betrachteten Analyse kann die Wanderung auch „umgekehrt“ für die Frage „Aus welchen Bundesländern stammen die behandelten Patienten?“ betrachtet werden. Die Prozentanteile aus Sicht der Behandlungsorte sind beispielsweise im Statistischen Jahrbuch 2004, S. 244f ⁹⁾ aufgeführt. Ebenso ist natürlich eine detailliertere Untersuchung, beispielsweise durch Verwendung kleinerer räumlicher Strukturen, denkbar.

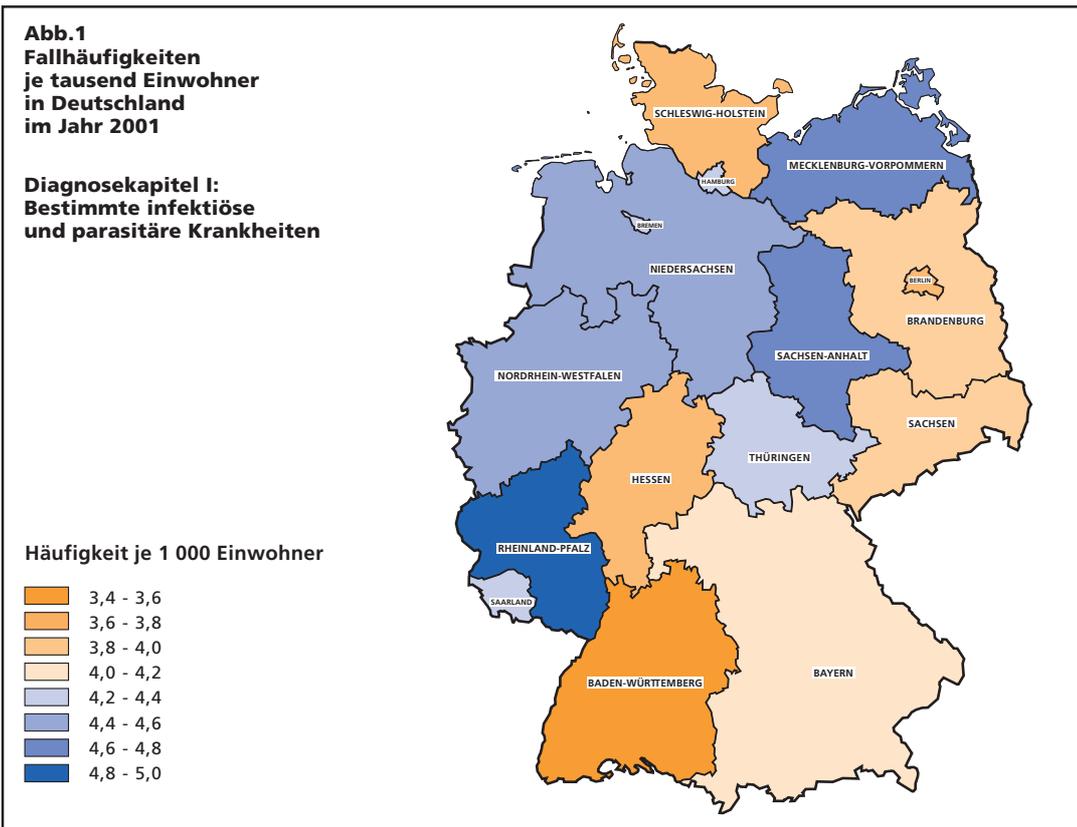
2. Verbleib von Patienten im Bundesland im Jahr 2001		
Wohnort (Bundesland)	Patient ist im Bundesland verblieben	
	ja	nein
Baden-Württemberg	95,2 %	4,8 %
Bayern	97,0 %	3,0 %
Berlin	96,2 %	3,8 %
Brandenburg	83,8 %	16,2 %
Bremen	91,1 %	8,9 %
Hamburg	90,3 %	9,7 %
Hessen	90,1 %	9,9 %
Mecklenburg-Vorpommern	95,0 %	5,0 %
Niedersachsen	88,7 %	11,3 %
Nordrhein-Westfalen	96,9 %	3,1 %
Rheinland-Pfalz	85,9 %	14,1 %
Saarland	92,7 %	7,3 %
Sachsen	95,8 %	4,2 %
Sachsen-Anhalt	92,2 %	7,8 %
Schleswig-Holstein	86,9 %	13,1 %
Thüringen	93,0 %	7,0 %

Quelle: Forschungsdatenzentrum der statistischen Ämter der Länder – Krankenhausstatistik, Teil II Diagnosen 2001

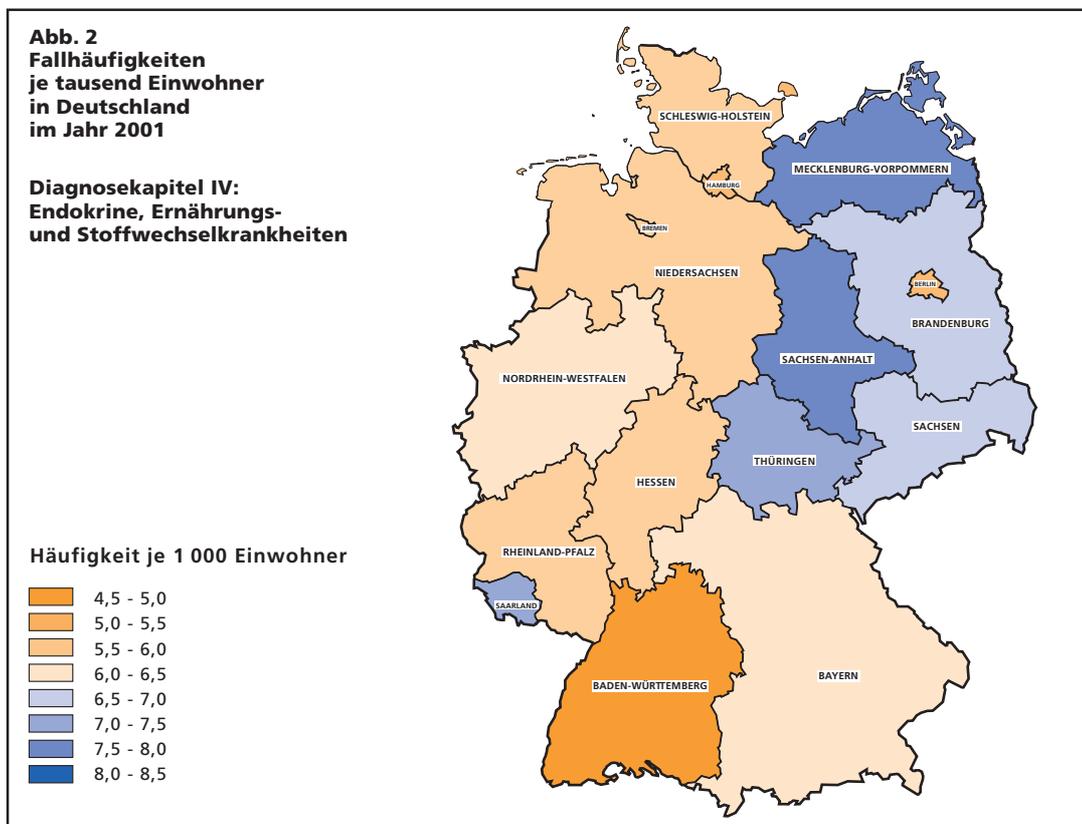
In einer weiteren Analyse werden die Häufigkeiten bestimmter Diagnosen nach dem Wohnort des Patienten bezogen auf die Einwohnerzahl berechnet und kartografisch dargestellt. Die Kartogramme in Abbildung 1 bis 3 stellen diese Verhältniszahlen in den einzelnen Bundesländern für drei ausgewählte Diagnosekapitel¹⁰⁾ dar. Die Klassengrenzen in diesen Abbildungen wurden dabei so gewählt, dass die Bandbreite der jeweiligen empirischen Häufigkeiten bestmöglich wiedergegeben wird.

Während es im Diagnosekapitel I (Bestimmte infektiöse und parasitäre Krankheiten) bundesweit keine gravierenden Unterschiede in der Behandlungshäufigkeit gibt, treten im Diagnosekapitel IV (Endokrine, Ernährungs- und Stoffwechselkrankheiten) in den neuen Bundesländern, gemessen an der Gesamtbevölkerung, deutlich mehr Krankenhausfälle auf als im übrigen Bundesgebiet. Bei den Psychischen und Verhaltensstörungen (Diagnosekapitel V) gibt es eher ein Süd-Nord-Gefälle. Baden Württemberg zeigt in allen drei untersuchten Kapiteln sehr geringe Behandlungshäufigkeiten.

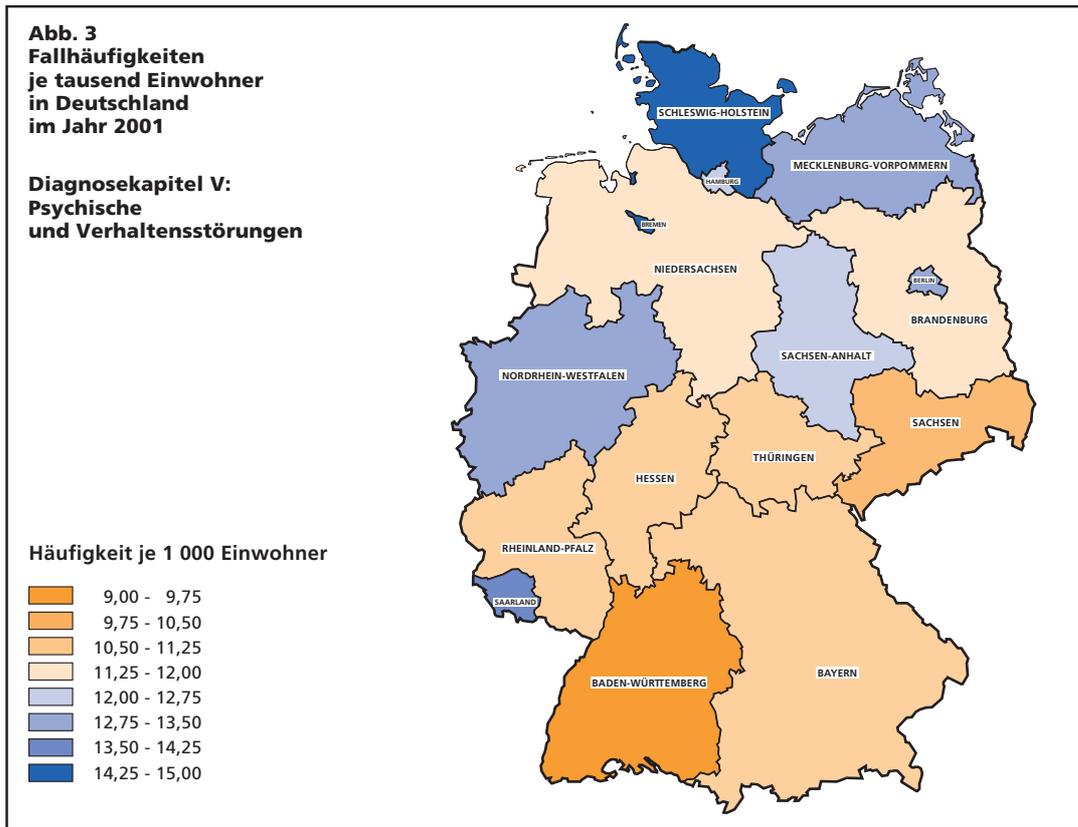
9) Alle Analysen im vorliegenden Artikel wurden einschließlich der Stundenfälle durchgeführt. Die Angaben im Statistischen Jahrbuch beziehen sich jedoch auf die Zahl der Behandlungen ohne Stundenfälle. – 10) Die Diagnosekapitel I bis XXI bilden die oberste Gliederungsstufe der ICD-10.



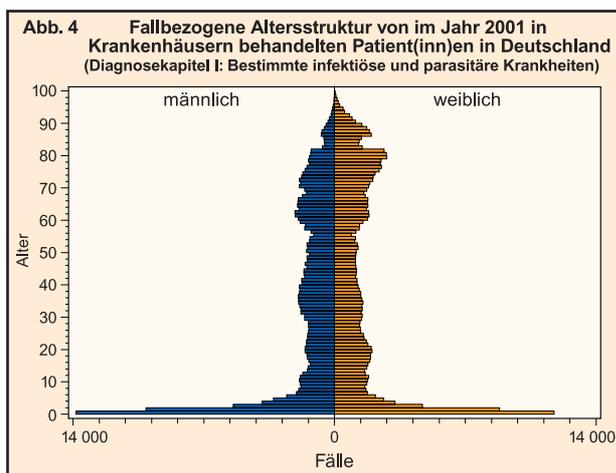
Quelle: Forschungsdatenzentrum der Statistischen Ämter der Länder – Krankenhausstatistik, Teil II Diagnosen 2001.



Quelle: Forschungsdatenzentrum der Statistischen Ämter der Länder – Krankenhausstatistik, Teil II Diagnosen 2001.



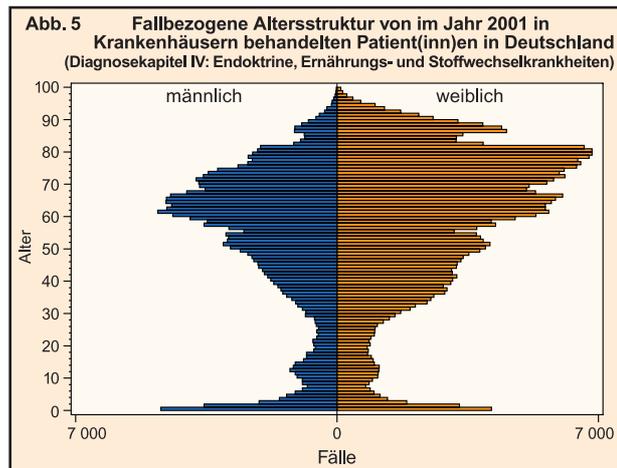
Die letzte hier vorgestellte Analyse befasst sich mit der Altersstruktur der behandelten Patienten.¹¹⁾ Dabei ist jedoch die erwähnte Fallbezogenheit der Diagnosestatistik zu berücksichtigen –



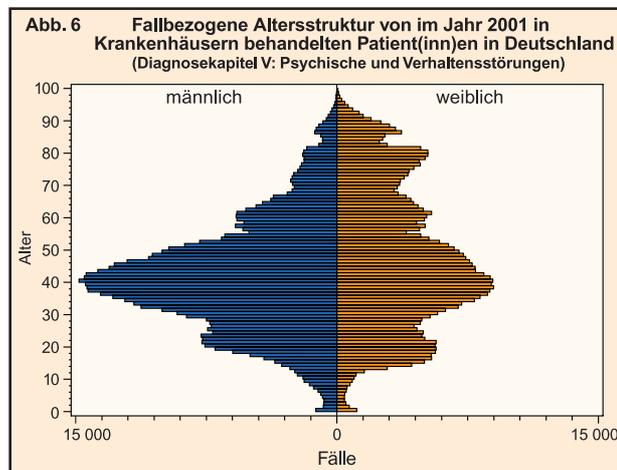
Patienten, welche im Jahr 2001 mehrfach in einem oder mehreren Krankenhäusern behandelt wurden, werden mehrfach gezählt. Daher sind die Alterspyramiden in den Abbildungen 4 bis 6 bezogen auf die Fälle, nicht aber auf die Patienten. Die Auswahl der Diagnosekapitel ist hierbei dieselbe wie bei der vorangegangenen Analyse.

Im Diagnosekapitel I ist eine Häufung für Säuglinge und Kleinkinder auffällig, da hier oft wegen Durchfallerkrankungen ein Krankenhausaufenthalt erforderlich wird. Bei älteren Patienten ist Wundrose eine häufige Einzeldiagnose.

11) Eine gleichartige, etwas detailliertere Analyse für 1999 aus Berliner Krankenhäuser entlassene Patienten ist in Bömermann (2001) zu finden.



Quelle: Forschungsdatenzentrum der Statistischen Ämter der Länder – Krankenhausstatistik, Teil II Diagnosen 2001.



Quelle: Forschungsdatenzentrum der Statistischen Ämter der Länder – Krankenhausstatistik, Teil II Diagnosen 2001.

Säuglinge und Kleinkinder werden auch im Diagnosekapitel IV überdurchschnittlich häufig behandelt, z. B. wegen Volumenmangels (Dehydratation). Bei den älteren Patienten spielen die Krankheitsklassen Diabetes mellitus und Krankheiten der Schilddrüse eine wichtige Rolle.

Bei der Häufung der Krankenhausfälle im Diagnosekapitel IV bei Männern um das 40. Lebensjahr ist die Einzeldiagnose Psychische und Verhaltensstörungen durch Alkohol überproportional häufig vertreten.

Die vorgestellten Analysen sind bewusst nicht allzu detailliert angelegt, um einer tiefergehenden Untersuchung der dargestellten Sachverhalte durch Nutzer des FDZ/L im Rahmen wissenschaftlicher Forschungsprojekte nicht vorzugreifen.

Über die verschiedenen Nutzungswege können selbstverständlich auch weitere Fragestellungen bezüglich der Krankenhaus-Diagnosestatistik untersucht werden, beispielsweise:

- Gibt es für bestimmte Diagnosen auffällige Altershäufungen? Was können deren Ursachen und Wirkungen sein?
- Gibt es besondere regionale Auffälligkeiten bestimmter Krankheiten (etwa Häufungen von Krebserkrankungen in der Nähe von Atomkraftwerken oder anderer Erkrankungen in Gebieten mit ärztlicher Unterversorgung)?
- Wie lang ist die mittlere Verweildauer bestimmter Patientengruppen (gegliedert etwa nach Wohnort, Alter oder Diagnose)?
- Welche Änderungen in der Diagnosestellung haben sich durch die Einführung der Abrechnung mittels DRG ergeben?

5 Fazit

Mit dem geplanten Scientific-Use-File stehen der Wissenschaft erstmals faktisch anonymisierte Mikrodaten der Krankenhaus-Diagnosestatistik zur Off-Site-Nutzung zur Verfügung. Auf Grund der Zusammenarbeit mit dem wissenschaftlichen Beraterkreis bei der Erstellung des Anonymisierungskonzeptes eignet es sich für einen großen Teil möglicher wissenschaftlicher Datenanalysen bezüglich der Diagnosen von Krankenhauspatienten. Für Mikrodaten der Krankenhaus-Diagnosestatistik stehen darüber hinaus als Zugangswege die kontrollierte Datenfernverarbeitung sowie die On-Site-Nutzung am Gastwissenschaftler-Arbeitsplatz zur Verfügung.

Im Forschungsprojekt zur beispielhaften Analyse der Diagnosedaten wurde das Wanderungsverhalten von Krankenhauspatienten und die relative Häufigkeit bestimmter Erkrankungen gegliedert auf Ebene der Bundesländer sowie die Altersstruktur der Patienten untersucht. Das Projekt soll einerseits das Interesse zur Nutzung von Daten der Diagnosestatistik für wissenschaftliche Zwecke fördern und eine Vorstellung von deren Analysepotenzial vermitteln. Andererseits können die Analysen Wissenschaftlern als Vorlage bei der Formulierung eigener Fragestellungen dienen.

Abkürzungsverzeichnis

AGS	–	Amtlicher Gemeindeschlüssel
BB	–	Brandenburg
BE	–	Berlin
BStatG	–	Bundesstatistikgesetz
BW	–	Baden-Württemberg
BY	–	Bayern
DIMDI	–	Deutsches Institut für medizinische Dokumentation und Information
DRG	–	Diagnosis Related Groups
DS I	–	Patient(inn)bezogener Datensatz des Scientific-Use-File
DS II	–	Krankenhausbezogener Datensatz des Scientific-Use-File
FDZ/L	–	Forschungsdatenzentrum der statistischen Landesämter
HB	–	Bremen
HE	–	Hessen
HH	–	Hamburg
ICD-10	–	10. Revision der Internationalen Klassifikation der Krankheiten
ICD-9	–	9. Revision der Internationalen Klassifikation der Krankheiten
MV	–	Mecklenburg-Vorpommern
NI	–	Niedersachsen
NW	–	Nordrhein-Westfalen
RP	–	Rheinland-Pfalz
SH	–	Schleswig-Holstein
SL	–	Saarland
SN	–	Sachsen
ST	–	Sachsen-Anhalt
TH	–	Thüringen

Literatur

Bömermann, H. (2001): Die Geschlechts- und Altersabhängigkeit des stationären Behandlungsaufkommens in den Berliner Krankenhäusern. *Berliner Statistik – Monatsschrift*, 8/01, 348 – 353.

DIMDI (Hg.) (2000): Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme (ICD-10), Version 2.0, Systematisches Verzeichnis. Kohlhammer Verlag: Stuttgart.

Scheffler, M. (2005): Ein Scientific-Use-File der Einzelhandelsstatistik 1999, *Wirtschaft und Statistik*, 3/2005, 197 – 200.

Statistisches Bundesamt (Hg.) (2004): Statistisches Jahrbuch 2004 für die Bundesrepublik Deutschland. Statistisches Bundesamt: Wiesbaden.

Zühlke, S., Hetke, U. (2003): Datenbedarf und Datenzugang: Ergebnisse der ersten Nutzerbefragung des Forschungsdatenzentrums der Statistischen Landesämter. *Allgemeines Statistisches Archiv*, Band 87, 321 – 334.

Zühlke, S., Zwick, M., Scharnhorst, S., Wende, T. (2004): „The research data centres of the Federal Statistical Office and the statistical offices of the Länder“, *Schmollers Jahrbuch*, Heft 4/2004, 567ff.

Autoren- und Kontaktangaben:

Gabriele Philipp, Statistisches Landesamt des Freistaates Sachsen, E-Mail: gabriele.philipp@statistik.sachsen.de

Dr. Olaf Schoffer, FDZ der Statistischen Ämter der Länder, Statistisches Landesamt des Freistaates Sachsen, E-Mail: olaf.schoffer@statistik.sachsen.de

Tarifbindung, betriebliche Lohnhöhe und Lohnstreuung im Produzierenden Gewerbe Baden-Württembergs

Stephan Bechtel, Wolf Dieter Heinbach, Harald Strotmann¹⁾

Ziel des vorliegenden Beitrags ist es, anhand von Mikrodaten aus der Verdienststrukturerhebung 2001 für das Produzierende Gewerbe Baden-Württembergs die Frage zu beantworten, welchen Einfluss kollektive Lohnfindungsregelungen auf die betriebliche Lohnhöhe und die betriebliche Lohnstreuung haben. Auf der Grundlage verschiedener multivariater Analysemethoden wird dabei gezeigt, dass ansonsten vergleichbare Betriebe, die der Tarifbindung unterliegen, ceteris paribus höhere Löhne zahlen als Betriebe mit individuellen Lohnvereinbarungen. Der Effekt der Tarifbindung ist dabei bei den Angestellten größer als bei den Arbeiter/-innen. Eine die Lohnstreuung reduzierende Wirkung der Tarifbindung kann für Arbeiter/-innen jedoch nicht für die Angestellten nachgewiesen werden.

1 Problemstellung und Ziele des Beitrags

Das System der Flächen- und Branchentarife in Deutschland steht seit geraumer Zeit verstärkt in der öffentlichen Diskussion (vgl. z. B. Artus 2001, Bispinck 2003 oder bereits früher Franz 1995). Neben dem Nutzen, den diese kollektiven Vereinbarungen für Betriebe und Arbeitnehmer/-innen aufweisen, werden dabei zunehmend auch die möglichen Schwachstellen und Nachteile von tarifvertraglichen Vereinbarungen beleuchtet. Kritiker sprechen daher teilweise auch bereits von der Gefahr einer „Erosion“ der Tarifvertragslandschaft, falls die Tarifparteien nicht selbst für eine rechtzeitige und hinreichende Reform derselben sorgen.

Auswertungen der Verdienststrukturerhebung der amtlichen Statistik, aber auch Berechnungen mit dem Betriebspanel des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) Nürnberg zeigen jedoch, dass die betriebliche Lohnfindung in Deutschland und im Südwesten weiterhin ganz entscheidend durch kollektive Lohnvereinbarungen geprägt ist (vgl. Statistisches Landesamt Baden-Württemberg 2004, Strotmann/Vogel 2004). Im Verarbeitenden Gewerbe Baden-Württembergs wurden im Jahr 2001 nach Auswertungen der Verdienststrukturerhebung rund 63 % der Vollzeitbeschäftigten unmittelbar auf der Grundlage kollektiver Lohnvereinbarungen entlohnt. 61 % aller Vollzeitbeschäftigten entfielen dabei auf Branchentarifverträge und 2 % auf Firmentarifvereinbarungen. Gleichzeitig muss jedoch konstatiert werden, dass tatsächlich in den vergangenen Jahren ein beträchtlicher Rückgang der Tarifbindung verzeichnet werden konnte, wurden doch 1995 noch knapp 75 % der Vollzeitbeschäftigten im Verarbeitenden Gewerbe und somit knapp 12 Prozentpunkte mehr tarifvertraglich entlohnt. Insofern zeigen die Ergebnisse auf, dass sich ein nicht unbeträchtlicher Teil der Betriebe – insbesondere kleinere Betriebe – im-

¹⁾ Verantwortlicher Autor: Dr. Harald Strotmann, Institut für Angewandte Wirtschaftsforschung (IAW) Tübingen, Ob dem Himmelreich 1, 72074 Tübingen, E-Mail: harald.strotmann@iaw.edu. Der Beitrag wurde im Rahmen eines Kooperationsprojektes mit dem Statistischen Landesamt Baden-Württemberg erstellt. Der herzliche Dank der Verfasser für die hervorragende Unterstützung gilt der unmittelbaren Projektbetreuerin Patricia Mödinger, Christel Much, den zuständigen Projektleitern des Statistischen Landesamtes Barbara Sinner-Bartels und Winfried Gruber sowie der Präsidentin Dr. Gisela Meister-Scheufelen.

mer stärker der Tarifbindung entzieht. Dem Bedürfnis eines Teils der Betriebe nach „etwas mehr Luft“ bei den Verhandlungen steht somit gleichzeitig eine beträchtliche Zahl der Betriebe gegenüber, welche die Vorteile der Tarifbindung weiterhin sieht und nicht aus der Tarifbindung ausgetreten ist.²⁾

Ein wesentlicher Vorteil von kollektiven Lohnvereinbarungen kann für die Betriebe darin bestehen, dass einzelne Arbeitgeber nicht in zeitaufwändige Lohnverhandlungen mit ihren Mitarbeiter/-innen eintreten müssen und somit Verhandlungs- und Transaktionskosten gespart werden können.³⁾ Darüber hinaus werden Konfliktpotenziale weitgehend aus den Betrieben ferngehalten, indem die Verhandlungen überbetrieblich stattfinden. Für die Arbeitnehmer stellen kollektive Lohnvereinbarungen eine gewisse Schutzfunktion vor einer zu großen Asymmetrie der „Verhandlungsmacht“ bei den Lohnfestsetzungen dar. Weiterhin soll eine schädliche Konkurrenz der Arbeitnehmer untereinander vermieden werden, indem Löhne und Arbeitsbedingungen durch Regelungen teilweise dem Wettbewerb entzogen werden.

In die Kritik geraten sind die Flächentarife und ihre Ausgestaltung in den vergangenen Jahren vor allem deshalb, weil in Frage gestellt wird, ob branchenweite Lohnabschlüsse in der bisherigen Form hinreichend Spielraum für betriebsindividuelle Differenzierungen lassen. In Zeiten einer zunehmenden Internationalisierung der Märkte und eines beschleunigten technologischen Wandels sei gerade die Möglichkeit der Betriebe, sich flexibel an veränderte Rahmenbedingungen anzupassen, eine lebensnotwendige Voraussetzung für die Wettbewerbsfähigkeit der Unternehmen und damit auch für die Sicherheit der betrieblichen Arbeitsplätze (vgl. z. B. Artus 2001).⁴⁾

Der vorliegende Beitrag untersucht empirisch anhand von Daten für das Produzierende Gewerbe in Baden-Württemberg, ob und in welchem Maße von der Tarifbindung ein Einfluss auf die betriebliche Lohnhöhe und die betriebliche Lohnstreuung ausgeht.⁵⁾ In Abschnitt 2 werden zunächst in sehr knapper Form theoretische Überlegungen angeführt, welchen Einfluss die Tarifbindung auf die Lohnstruktur, im Speziellen auf die betriebliche Lohnhöhe und die betriebliche Lohnstreuung haben könnte. Abschnitt 3 stellt dann die Verdienststrukturerhebung, die für die weiteren empirischen Analysen verwendete Teilstichprobe sowie die Operationalisierung des verwendeten Lohnbegriffes vor, bevor in Abschnitt 4 die empirischen Analysen erläutert und interpretiert werden. Abschnitt 5 fasst zentrale Ergebnisse zusammen, diskutiert methodische Probleme und gibt einen Ausblick auf weiteren Forschungsbedarf.

2) Arndt (2004) findet für Baden-Württemberg anhand des IAB-Betriebspanels einen erheblichen Anteil der Betriebe, die sich in jedem Jahr neu für eine kollektive Lohnfindung entscheiden. Darüber hinaus ist auch der Anteil der Betriebe, die sich nach Verlassen der Tarifbindung bereits wenig später wieder in die Bindung an einen Flächentarifvertrag begeben, durchaus erheblich. – 3) Vgl. hierzu z. B. Kohaut/Schnabel (2003), S. 319 ff. – 4) Solche Anpassungen an die betriebliche Ebene können durch Öffnungs- und Härtefallklauseln im Flächentarifvertrag erlaubt werden. In der Vergangenheit wurden in zahlreichen Tarifbereichen vor allem Öffnungsklauseln, die gerade zur Wettbewerbsverbesserung und zur Sicherung der Beschäftigung angewendet werden können, in den Tarifverträgen verankert. Je nach Art und Ausgestaltung solcher Klauseln ist das Ausmaß der Öffnung gemessen am jeweils betroffenen Beschäftigtenanteil beträchtlich (vgl. Heinbach 2005). Trotz der bestehenden Öffnungsmöglichkeiten ist weiterhin umstritten, ob bereits ein ausreichendes Maß an Öffnung erreicht wurde oder ob ein zusätzlicher Bedarf an weiteren Öffnungsklauseln besteht (vgl. Bispinck 2003). Die Analyse der Verbreitung und Wirkungen von Öffnungsklauseln ist nicht Thema dieses Beitrags, wird jedoch im Rahmen eines laufenden IAW-Projekts im Auftrag der Deutschen Forschungsgemeinschaft (DFG) untersucht. Vgl. dazu auch die Ausführungen im zusammenfassenden Abschnitt 5. – 5) Ähnliche Untersuchungen wurden in Deutschland bislang nur in Niedersachsen von Gerlach/Stephan durchgeführt (z. B. Gerlach/Stephan 2005, 2002), an die sich die vorliegende Studie für Baden-Württemberg anlehnt.

2 Hypothesen über Tarifbindung und Lohnstruktur

Derzeit existiert noch kein geschlossener theoretischer Modellrahmen, der dem Lohnbildungsmechanismus und den institutionellen Rahmenbedingungen in Deutschland gerecht wird (vgl. Franz 2003, S. 289). Insofern beleuchten die weiteren theoretischen Überlegungen einzelne Aspekte der möglichen Bedeutung der Tarifbindung für die betriebliche Lohnhöhe und die Lohnstreuung. Grundsätzlich unterstellen sämtliche Modelle der Lohnfindung, dass Unternehmen ihren Gewinn maximieren, während Gewerkschaften versuchen, den Nutzen ihrer Mitglieder/-innen zu maximieren, der einerseits durch die Lohnhöhe und andererseits durch die Beschäftigung bestimmt wird (vgl. z. B. Franz 2003).

Betrachtet man mögliche Zusammenhänge zwischen Tarifbindung und *betrieblicher Lohnhöhe*, so könnte man aus theoretischer Sicht zunächst erwarten, dass tarifgebundene Betriebe unter sonst gleichen Bedingungen höhere Löhne zahlen, da die größere Verhandlungsmacht der kollektiven Organisation der Interessen auf der Arbeitnehmerseite zu besseren Lohnabschlüssen für die Beschäftigten führt. Auch Überlegungen auf der Grundlage der Insider-Outsider-Theorie lassen vermuten, dass die Gewerkschaften in der Lage sind, für die beschäftigten Insider möglichst gute Lohnabschlüsse zu verhandeln. Darüber hinaus sind die Arbeitgeber selbst bereit, gewisse Lohnzugeständnisse zu machen, da sie durch die kollektiven Vereinbarungen Verhandlungs- und Konfliktkosten sparen, sodass höhere Löhne durchaus den Nutzen der Unternehmen vergrößern können. Auch ist der Doppelcharakter der Entlohnung von Mitarbeitenden zu berücksichtigen, da Löhne einerseits Kostenfaktor sind, andererseits jedoch eine höhere Entlohnung auch die Produktivität der Mitarbeiter/-innen anregt (vgl. Freeman/Medoff 1984) und mögliche Fluktuations- und Suchkosten senken kann. Diese Argumentation gilt grundsätzlich sowohl für Arbeiter/-innen als auch für Angestellte. Gegen eine höhere Lohnhöhe in tarifgebundenen Betrieben könnte zum Beispiel sprechen, dass einige Tarifverträge inzwischen Öffnungsmöglichkeiten vorsehen, die den Lohndruck nach oben zumindest abschwächen könnten.

Hinsichtlich der *betrieblichen Lohnstreuung* sind die Hypothesen etwas weniger eindeutig (vgl. u. a. Freeman 1982, Blau/Kahn 1999, Lemieux 1998).⁶⁾ So könnte man einerseits erwarten, dass in Betrieben mit Tarifbindung die Streuung sowohl innerhalb als auch zwischen den Betrieben geringer ausfällt. Begründet werden kann dies zunächst durch die Idee gleichen Lohns für gleiche Arbeit, die für die gewerkschaftliche Lohnpolitik eine zentrale Maxime darstellt, sowie durch ergänzende Gerechtigkeits- oder Solidaritätsziele der Arbeitnehmervertreter. Auch könnte man vermuten, dass in tarifgebundenen Betrieben für die Vorgesetzten ein geringerer Spielraum zur Lohndiskriminierung besteht. Geht man realistischerweise von einer linkssteilen Einkommensverteilung aus, so wird das Einkommen des Medianmitglieds einer Gewerkschaft geringer ausfallen als das durchschnittliche Einkommen, wodurch ein Lohndruck nach oben am unteren Ende der Lohnverteilung entsteht. Andererseits könnten mögliche Gegenargumente, die gegen eine geringere Lohnstreuung in tariflichen Betrieben sprechen, darin bestehen, dass auch tarifgebundene Betriebe verstärkt Entlohnungsformen einsetzen, die stärker an der Produktivität der Mitarbeiter/-innen orientiert sind.⁷⁾ Darüber hinaus könnte bei einer zu gleichen

6) Etwas ausführlicher ist die zusammenfassende Darstellung verschiedener Argumente bei Gerlach/Stephan (2005). – 7) Allerdings zeigen Auswertungen des IAB-Betriebspanels für Westdeutschland, dass sich tarifgebundene Betriebe *ceteris paribus* mit geringerer Wahrscheinlichkeit für die Einführung von erfolgsabhängigen Entlohnungsmodellen entscheiden (vgl. Strotmann 2005).

Entlohnung die Gefahr bestehen, dass die höher Qualifizierten ihren Einkommensanstieg als zu gering empfinden und daher die Politik der Gewerkschaften nicht mehr unterstützen (vgl. Acemoglu et al. 2001). Zumal argumentiert wird, dass kollektive Verträge gerade vor dem Hintergrund der erheblichen organisatorischen Veränderungsprozesse, die immer stärker allgemeine und unterschiedliche Fähigkeiten statt Spezialisierung erfordern, ineffizient seien (vgl. Lindbeck/Snower 2001).

3 Zur Datengrundlage: Verdienststrukturerhebung 2001

Um die aufgeworfenen Hypothesen zu testen wird die Verdienststrukturerhebung 2001 des Statistischen Landesamtes Baden-Württemberg verwendet. Dabei handelt es sich um einen Employer-Employee-Datensatz (vgl. dazu auch Abowd/Kramarz 1999), der gleichzeitig sowohl Informationen über die Betriebe als auch deren Beschäftigte enthält.⁸⁾ Der Datensatz basiert auf einer zweistufigen, repräsentativen Stichprobe von rund 3 000 Betrieben in Baden-Württemberg mit 10 oder mehr Beschäftigten, aus denen eine Stichprobe von insgesamt rund 110 000 Einzeldatensätzen von Arbeitnehmerinnen und Arbeitnehmern gezogen wurde. Abgedeckt sind das Produzierende Gewerbe sowie ausgewählte Bereiche des Dienstleistungssektors. Für die weiteren Auswertungen werden jedoch ausschließlich Betriebe des Produzierenden Gewerbes betrachtet. Darüber hinaus wird die Abschneidegrenze auf Betriebe mit 100 oder mehr Beschäftigten erhöht, um eine hinreichende betriebsinterne Streuung der Löhne zu ermöglichen. Da Mitarbeiter/-innen in Führungspositionen häufig außertariflich entlohnt werden und sehr heterogene Lohnstrukturen aufweisen, werden Angestellte der Führungsebenen mit Aufsicht- oder Dispositionsbefugnis und mit eingeschränkter Aufsicht- und Dispositionsbefugnis aus den Analysen ausgeklammert. Verzerrungen durch Teilzeitarbeit und geringfügige Beschäftigung werden ausgeschlossen, indem nur Personen betrachtet werden, die eine vereinbarte Wochenarbeitszeit von 30 Stunden oder mehr aufweisen und weder teilzeitbeschäftigt noch in Ausbildung sind. Nach den Bereinigungsmaßnahmen verbleiben in der Stichprobe Angaben zu rund 19 000 Arbeiterinnen und Arbeitern sowie 10 000 Angestellten, die sich auf mehr als 600 Betriebe verteilen.

Die Tarifbindung eines Betriebes bzw. seiner Beschäftigten kann anhand des Tarifvertragschlüssels identifiziert werden, wobei man eine Entlohnung nach Flächentarifvertrag von einer Entlohnung nach Firmentarifvertrag und gemäß individueller Lohnvereinbarung unterscheiden kann.⁹⁾

Als betrieblicher Lohn werden Bruttostundenlöhne im Rahmen der regulären, vertraglich vereinbarten Arbeitszeit betrachtet. Sowohl im Lohn als auch bei der Arbeitszeit wird daher jeweils bezahlte Mehrarbeit herausgerechnet. Implizit wird somit unterstellt, dass keine unbezahlte Mehrarbeit stattfindet bzw. diese keinen systematisch unterschiedlichen Einfluss auf die Lohnhöhe in Betrieben mit und ohne Tarifbindung hat. Die Analysen erfolgen dabei getrennt nach Arbeiter/-innen und Angestellten.

8) Zur Verdienststrukturerhebung im Detail vgl. den Beitrag von Hafner (2006) in diesem Band. – 9) In denjenigen Betrieben, in denen ein Teil der beschäftigten Arbeitnehmer/-innen durch einen Flächentarifvertrag und der andere Teil durch individuelle Vereinbarung oder Firmentarifvertrag entlohnt werden, wurden die Betriebe nach dem Schwerpunktprinzip demjenigen Verhandlungsregime zugeordnet, das für die absolut häufigste Zahl der Arbeitnehmer/-innen Anwendung fand.

4 Tarifbindung, Lohnhöhe und Lohnstreuung

4.1 Ausgangspunkt: Deskriptiver Befund

Einfache Kerndichteschätzungen der Verteilungen der logarithmierten Bruttostundenlöhne für Arbeiter/-innen und Angestellte scheinen sowohl die Hypothese höherer Löhne als auch einer geringeren Lohnstreuung in tarifgebundenen Betrieben zu unterstreichen. Sowohl die Medianlöhne als auch das erste und dritte Quartil fallen in Betrieben mit individuellen Lohnvereinbarungen beträchtlich geringer aus als in Betrieben mit Flächen- oder Haustarifvertrag. Der durchschnittliche logarithmierte Bruttostundenlohn der Arbeiter/-innen beträgt in den tarifgebundenen Betrieben 2,74, was einem Bruttostundenlohn von knapp 16 Euro entspricht.¹⁰⁾ In den Betrieben mit individuellen Lohnvereinbarungen liegt er gut 7 % niedriger bei rund 2,54 bzw. gut 13 Euro. Für die Angestellten erhält man ein ähnliches Bild, einem Bruttostundenlohn von rund 20 Euro steht ein Lohn von knapp 17 Euro in Betrieben mit individuellen Vereinbarungen gegenüber. Doch ist an dieser Stelle die Schlussfolgerung nicht zulässig, dass sich die höheren Löhne auch ursächlich auf die Tarifbindung zurückführen lassen, da sich Betriebe mit und ohne Tarifbindung auch hinsichtlich zahlreicher weiterer möglicher Einflussfaktoren, wie z. B. der Qualifikationsstruktur der Mitarbeiter/-innen oder der Branchenstruktur, systematisch unterscheiden können. Eine Beantwortung der Frage, ob und in welchem Maße gerade die Tarifbindung zu höheren Löhnen führt, kann erst im Rahmen multivariater Analysen geeignet untersucht werden, obwohl auch dort einige Probleme verbleiben.

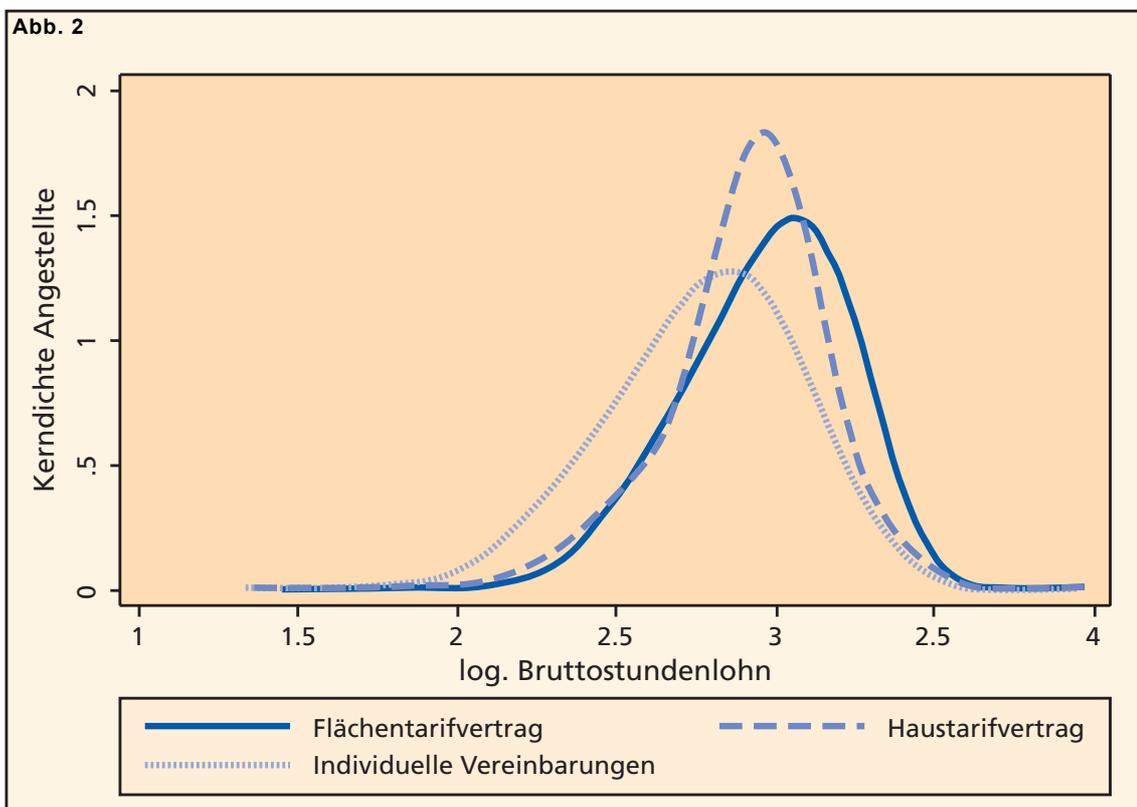
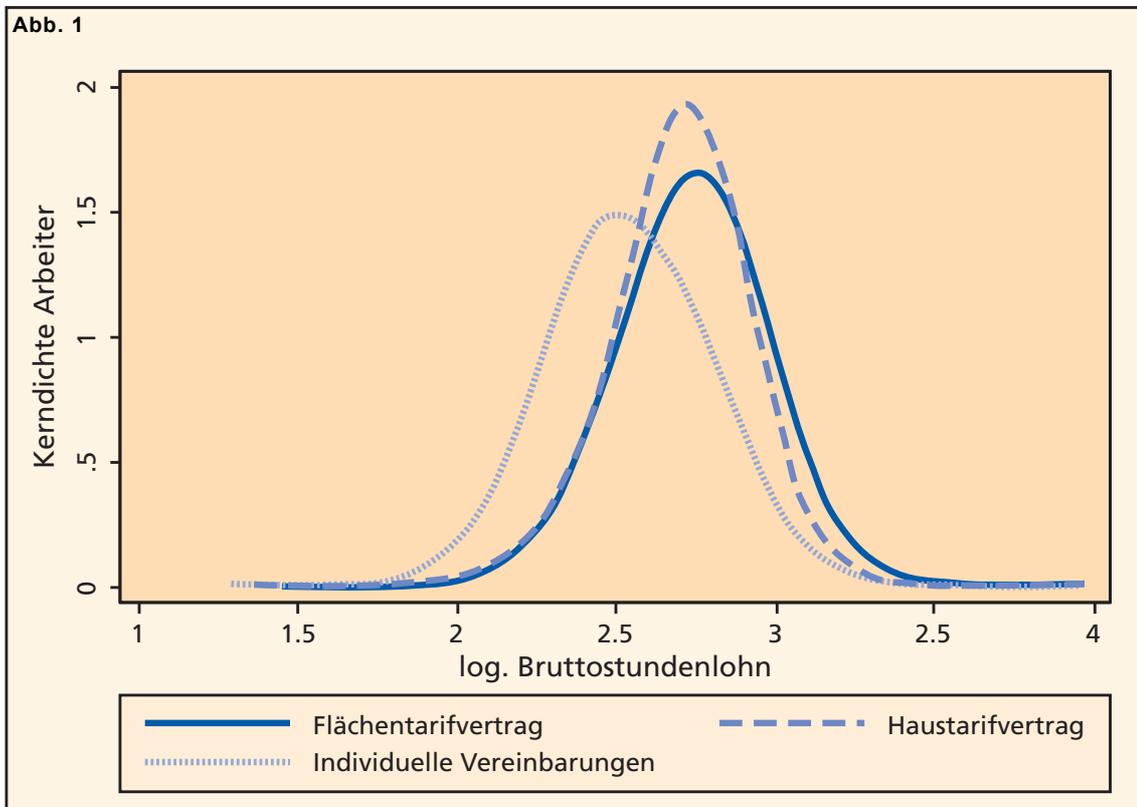
Die Streuung der Löhne ist ebenfalls bei deskriptiver Betrachtung sowohl bei den Arbeiter/-innen als auch bei den Angestellten in tarifgebundenen Betrieben geringer. Während die Standardabweichung der logarithmierten Bruttostundenlöhne bei den Arbeiter/-innen 0,23 und in Betrieben mit Haustarifverträgen 0,19 beträgt, liegt die entsprechende Standardabweichung in Betrieben mit individuellen Lohnvereinbarungen bei 0,25. Ein ähnliches Bild findet man auch hier bei den Angestellten. Auch bei Betrachtung der Variationskoeffizienten, bei denen die Standard-

1. Lohnhöhe und Lohnstreuung in Betrieben mit Flächentarifvertrag (FTV), Firmentarifvertrag (HTV) und individuellen Lohnvereinbarungen (IV) im Vergleich, getrennt für Arbeiter/-innen und Angestellte						
Merkmal	Gewerbliche Arbeitnehmer/-innen			Angestellte		
	FTV	HTV	IV	FTV	HTV	IV
Nominaler Bruttostundenlohn in Euro Mittelwert	15,84	15,20	13,10	20,01	18,85	16,81
Logarithmierter Bruttostundenlohn in Euro Mittelwert	2,74	2,70	2,54	2,96	2,91	2,78
Standardabweichung	0,23	0,19	0,25	0,26	0,24	0,30
Variationskoeffizient	0,084	0,071	0,097	0,088	0,082	0,109
Varianzanalyse						
– Standardabweichung innerhalb der Betriebe	0,156	0,154	0,178	0,217	0,229	0,232
– Standardabweichung zwischen den Betrieben	0,169	0,120	0,173	0,145	0,073	0,195
Anzahl Beobachtungen	14 263	817	4 027	7 444	378	2 092
Anzahl Betriebe	474	23	150	479	24	165
Anzahl Flächentarifverträge	69			63		

Quelle: Verdienststrukturerhebung Baden-Württemberg 2001, IAW-Berechnungen

10) Dabei handelt es sich um ungewichtete Größen, da die weiteren Regressionen ebenfalls ungewichtet durchgeführt werden. Allerdings spielt die Gewichtung letztlich für die inhaltlichen Ergebnisse keine wesentliche Rolle.

Lohnhöhe und Lohnstreuung in Betrieben mit Flächentarifvertrag, Firmentarifvertrag und individuellen Lohnvereinbarungen im Vergleich, Kerndichteschätzungen getrennt für Arbeiter/-innen und Angestellte



abweichung auf den Mittelwert bezogen und somit die Streuung als relatives Phänomen betrachtet wird, bestätigt sich die höhere Lohnstreuung in Betrieben mit individuellen Lohnvereinbarungen. Die Lohnstreuung bei den gewerblichen Arbeitnehmer/-innen ist dabei größer als die entsprechende Lohnstreuung bei den Angestellten.

Die höhere Lohnstreuung in Betrieben mit individuellen Lohnvereinbarungen resultiert dabei sowohl aus einer höheren Streuung der Löhne innerhalb der Betriebe als auch aus größeren Lohnunterschieden zwischen den Betrieben (vgl. Tabelle 1). Während bei den Arbeiter/-innen dabei die Lohnstreuungen innerhalb und zwischen den Betrieben ähnlich groß ausfallen, variieren die Bruttostundenlöhne der Angestellten innerhalb der Betriebe erheblich stärker als zwischen den Betrieben.

Auch für den deskriptiven Streuungsbefund gilt jedoch, dass er alleine noch keine fundierten Aussagen darüber zulässt, ob und in welchem Maße die Tarifbindung tatsächlich ursächlich für die geringere Lohnstreuung ist. Um zu besseren Aussagen zu gelangen, sind multivariate Analysen erforderlich, bei denen andere mögliche Einflussgrößen der betrieblichen Lohnhöhe und der betrieblichen Lohnstreuung gleichzeitig kontrolliert werden. In Abschnitt 4.2 werden zunächst die empirischen Ergebnisse zu den Determinanten der betrieblichen Lohnhöhe vorgestellt, bevor sich Abschnitt 4.3 einer entsprechenden Ursachenanalyse für die betriebliche Lohnstreuung widmet.

4.2 Tarifbindung und betriebliche Lohnhöhe

Im Rahmen der weiteren Analysen soll eine Lohnfunktion in der Tradition von Mincer (1974) geschätzt werden, indem der logarithmierte Bruttostundenlohn des i -ten Betriebs $\ln w_i$ durch einen Vektor x_i exogener Variablen erklärt wird.

$$(1) \quad \ln w_i = x_i' \beta + \varepsilon_i$$

β ist der zu schätzende Koeffizientenvektor und ε_i ein annahmegemäß identisch und unabhängig verteilter Zufallsterm, der sämtliche Einflüsse auffangen soll, die nicht explizit in dem Modell berücksichtigt werden.

Der semilogarithmische Ansatz ermöglicht es, die geschätzten Koeffizienten näherungsweise als prozentuale Effekte gegenüber der Referenzkategorie zu interpretieren. Da die deskriptive Betrachtung der Verteilung der logarithmierten Bruttostundenlöhne gezeigt hat, dass gerade nach oben teilweise erhebliche Ausreißer vorliegen, wird der Koeffizientenvektor nicht nur mit Hilfe einer Ordinary-Least-Squares (OLS)-Schätzung (mit heteroskedastizitätskonsistenten Standardfehlern, vgl. White 1980) geschätzt, sondern alternativ auch mit Hilfe zweier gegenüber Ausreißern robusterer Methoden. Einerseits wird ergänzend eine Medianregression durchgeführt, bei der letztlich die Summe der absoluten Fehler und nicht – wie bei OLS – die Summe der quadrierten Fehler minimiert wird. Andererseits werden robuste Regressionen durchgeführt, bei denen extreme Werte in ihrer Bedeutung für die Regressionsgleichung heruntergewichtet, im Extremfall sogar ausgeschlossen, werden (vgl. Rousseeuw/Leroy 1987).

Die Verdienststrukturerhebung ermöglicht es, als Employer-Employee-Datensatz im Vektor x_i sowohl individuelle als auch betriebliche Einflussgrößen der Entlohnung zu berücksichtigen. Da im Rahmen der folgenden Analysen die Lohnhöhe auf betrieblicher Ebene untersucht wird, werden auch individuelle Merkmale jeweils auf die betriebliche Ebene bezogen. In diesem Beitrag wird kein umfassender theoretischer Überblick über mögliche Determinanten der betrieblichen Lohnhöhe gegeben, sondern nur sehr knapp erläutert, welche Einflussfaktoren im Rahmen der Modelle berücksichtigt werden. Im Einzelnen werden folgende Variablen einbezogen:

Um die für den Beitrag zentrale Frage, den möglichen Einfluss der Tarifbindung auf die betriebliche Lohnhöhe, zu überprüfen, werden Dummyvariablen für Betriebe mit Branchen- bzw. Haustarifvertrag berücksichtigt, als Referenzkategorien fungieren Betriebe mit individuellen Lohnvereinbarungen.

Die Humankapitaltheorie in der Tradition von Becker (1975) lässt erwarten, dass mit zunehmender Qualifikation der Beschäftigten deren Entlohnung ansteigt. Daher werden Kontrollvariablen für das *allgemeine Humankapital* und das *firmenspezifische Humankapital* der in einem Betrieb Beschäftigten berücksichtigt. Als erklärende Variable wird daher zunächst die durchschnittliche Dauer der schulischen und beruflichen Ausbildung der Beschäftigten einbezogen. Darüber hinaus werden auf der Grundlage der „Leistungsgruppen“ Variablen für die Anteile der in den verschiedenen Qualifikationsgruppen Beschäftigten gebildet. Für die Arbeiter/-innen werden angeleitete Arbeiter/-innen, Facharbeiter/-innen und qualifizierte Facharbeiter/-innen berücksichtigt, wobei der Anteil der ungelerten Arbeiter/-innen als Referenzkategorie dient. Angestellte mit einfacher Tätigkeit bilden für die berücksichtigten Anteile von Angestellten ohne Entscheidungsbefugnis, Angestellten mit selbstständiger Tätigkeit und solchen mit besonderen Fachkenntnissen die Referenzgröße. Die mittlere Betriebszugehörigkeitsdauer der Beschäftigten informiert als Indikatorvariable über die Akkumulation von firmenspezifischem Humankapital. Das durchschnittliche Alter der Beschäftigten (und ggf. quadriert) wird als Indikator für eine mögliche *Seniorentlohnung* der Beschäftigten in die Modelle eingebaut.

Da die Höhe der Entlohnung entscheidend davon abhängt, ob Sonntags- oder Nachtarbeit anfallen, für die jeweils Zulagen geleistet werden, wird diesem Aspekt mittels zweier Dummyvariablen Rechnung getragen. Im gewerblichen Bereich ist die Form der Entlohnung (Mischlohn, Akkordlohn, Prämienlohn) von wesentlicher Bedeutung für die Lohnhöhe. Daher werden jeweils die Anteile der Beschäftigten, die über diese Formen der Entlohnung verfügen, als erklärende Variablen berücksichtigt, wobei der Zeitlohn als Referenzkategorie fungiert. Für die Angestellten wird eine Variable einbezogen, ob eine Entlohnung für eine technische Tätigkeit vorliegt oder nicht.

Der *Frauenanteil* der Beschäftigten eines Betriebes wird ebenfalls als erklärende Variable eingebaut, um zu überprüfen, ob sich auf betrieblicher Ebene Indizien für eine Lohndiskriminierung von Frauen nachweisen lassen.

Aus verschiedenen Studien (z. B. Brown et al. 1990) und den Ausführungen in Kapitel 1 dieser Studie ist bekannt, dass auch die *Betriebsgröße* eine wesentliche Rolle bei der Erklärung der Lohnhöhe spielen kann, da größere Betriebe häufig in der Lage sind, aufgrund ihrer Produktivitätsvorteile ihren Beschäftigten auch höhere Löhne zu zahlen. Die Größe der Betriebe wird im Weiteren anhand der Beschäftigtenzahl der Betriebe operationalisiert und in Form von Dummyvariablen für verschiedene Größenklassen berücksichtigt.

Die Bedeutung des *Branchenumfelds*, in dem ein Betrieb agiert, kann unter sonst gleichen Bedingungen ebenfalls einen prägenden Einfluss auf die Lohnhöhe haben. Zwar werden die Wettbewerbsbedingungen in den einzelnen Branchen hier nicht detailliert abgebildet, Dummyvariablen für die Zweisteller der Wirtschaftsklassifikation WZ 93 sorgen jedoch dafür, dass Brancheneinflüsse auf die Lohnhöhe zumindest in allgemeiner Form in die Modellschätzungen einbezogen werden.

Zur Überprüfung der Stabilität der Ergebnisse wurden sowohl für Arbeiter/-innen als auch für Angestellte eine Reihe unterschiedlicher Modelle geschätzt, die sich insbesondere danach unterscheiden, ob sie mit Hilfe von quadrierten Termen einen nichtlinearen Einfluss der Variablen Alter und Betriebszugehörigkeitsdauer abbilden. Außerdem wurde überprüft, inwieweit die zwischen den eben genannten Variablen bestehende Korrelation über die Hinzunahme eines Interaktionseffekts berücksichtigt werden muss. Tabelle 2.1 und 2.2 geben die Ergebnisse der Schätzungen für Arbeiter/-innen und Angestellte wieder.

Das zentrale Ergebnis hinsichtlich der Bedeutung der Tarifbindung für die betriebliche Lohnhöhe von Arbeiterinnen und Arbeitern sowie Angestellten ist, dass ansonsten vergleichbare Betriebe mit Branchentarifvertrag auch bei Kontrolle weiterer Einflussgrößen sowie unabhängig von der Modellspezifikation und der gewählten Schätzmethode einen statistisch signifikant höheren Bruttostundenlohn an Arbeiter/-innen und an Angestellte zahlen als Betriebe mit individuellen Lohnvereinbarungen. Der durch die Tarifbindung entstehende Lohnvorteil beträgt dabei je nach Schätzung für die Arbeiter/-innen zwischen 6 und 7 %. Auch Betriebe mit Haustarifvertrag zahlen statistisch signifikant höhere Löhne als vergleichbare Betriebe mit individuellen Lohnvereinbarungen, wobei der Lohnvorteil rund 4 bis 7 % beträgt. Bei den Angestellten beträgt der relative Lohnvorsprung von tarifgebundenen Betrieben gegenüber ansonsten vergleichbaren Betrieben ohne Tarifbindung rund 13 bis 14 % und ist somit größer als bei den Arbeiterinnen und Arbeitern.

Für die weiteren Einflussvariablen ergeben sich weitgehend die theoretisch erwarteten und somit sehr plausiblen Ergebnisse. Qualifikation ist den Voraussagen der Humankapitaltheorie folgend ein wesentlicher Faktor für die Erklärung der betrieblichen Lohnhöhe. Dies wird sowohl durch die Variable für die Dauer der schulischen und beruflichen Ausbildung als auch durch die Variablen für die Anteile der verschiedenen qualifizierten Mitarbeiter/-innen bei den Arbeiter/-innen und Angestellten bestätigt. Der Verdienstabstand zur Gruppe der ungelerten Arbeiter/-innen steigt dabei erwartungsgemäß von Leistungsgruppe zu Leistungsgruppe an.¹¹⁾

Die Betriebszugehörigkeitsdauer der Mitarbeiter/-innen weist sowohl für Arbeiter/-innen als auch für Angestellte einen positiven Einfluss auf die betriebliche Lohnhöhe auf, der jedoch – wie der signifikant negative Koeffizient des quadrierten Terms für die Löhne der Arbeiter/-innen zeigt – für diese mit zunehmender Dauer der Betriebszugehörigkeit an Stärke verliert. Dies spricht für das Vorliegen von Senioritätseffekten bei der betrieblichen Entlohnung. Das Alter spielt nur bei der betrieblichen Lohnhöhe im Bereich der Angestellten eine statistisch signifikante Rolle, bei den gewerblichen Arbeitnehmern dagegen besteht kein signifikanter Einfluss des durchschnittlichen Alters der Arbeiter/-innen auf die betriebliche Lohnhöhe.¹²⁾

11) Die in den Tabellen 2.1 und 3.1 ausgewiesenen Qualifikationsniveaus „angelernter Arbeiter/-innen“, „Facharbeiter/-innen“ und „qualifizierter Facharbeiter/-innen“ stehen für die Leistungsgruppen 2, 1 und 0. Referenz ist die Leistungsgruppe 3, die ungelerten Arbeiter/-innen beinhaltet. – 12) Dieses Resultat deckt sich mit deskriptiven Auswertungen zur gleichen Thematik, vgl. z. B. Mödinger (2003), S. 8.

2.1 Determinanten der betrieblichen Lohnhöhe, Ergebnisse von OLS-Regressionen, Robusten Regressionen und Medianregressionen für Arbeiter/-innen			
Merkmal	Arbeiter/-innen		
	OLS	Robuste Regression	Medianregression
Alter/10	-0,137 (0,341)	0,038 (0,725)	-0,085 (0,382)
(Alter/10) ²	0,016 (0,336)	-0,004 (0,759)	0,011 (0,347)
Betriebszugehörigkeitsdauer/10	0,187 (0,000)**	0,141 (0,000)**	0,15 (0,000)**
(Betriebszugehörigkeitsdauer/10) ²	-0,037 (0,006)**	-0,024 (0,041)*	-0,028 (0,007)**
Jahre schul./berufl. Ausbildung	0,041 (0,003)**	0,041 (0,001)**	0,042 (0,000)**
Frauenanteil	-0,152 (0,000)**	-0,174 (0,000)**	-0,153 (0,000)**
Ungelernte Arbeiter/-innen (Referenzkategorie)	-	-	-
Anteil qualifizierter Facharbeiter/-innen	0,264 (0,000)**	0,232 (0,000)**	0,231 (0,000)**
Anteil Facharbeiter/-innen	0,224 (0,000)**	0,19 (0,000)**	0,203 (0,000)**
Anteil angelernter Arbeiter/-innen	0,102 (0,000)**	0,075 (0,001)**	0,094 (0,000)**
Zeitlohn (Referenzkategorie)	-	-	-
Prämienlohn	0,044 (0,040)*	0,054 (0,006)**	0,071 (0,000)**
Akkordlohn	0,094 (0,003)**	0,123 (0,000)**	0,151 (0,000)**
Prämien- und Akkordlohn	0,165 (0,000)**	0,145 (0,09)	0,157 (0,010)*
Mischlohn	0,018 (0,564)	0,02 (0,426)	0,012 (0,582)
Sonntagsarbeit	0,088 (0,001)**	0,053 (0,008)**	0,05 (0,004)**
Nachtarbeit	0,033 (0,118)	0,054 (0,000)**	0,063 (0,000)**
Betriebsgröße 1 – 199 (Referenzkategorie)	-	-	-
Betriebsgröße 200 – 499	0,031 (0,006)**	0,029 (0,008)**	0,037 (0,000)**
Betriebsgröße 500 – 999	0,032 (0,041)*	0,037 (0,010)*	0,05 (0,000)**
Betriebsgröße 1 000 und mehr	0,073 (0,000)**	0,072 (0,000)**	0,074 (0,000)**
Individuelle Lohnvereinbarung (Referenzkategorie)	-	-	-
Flächentarifvertrag	0,059 (0,000)**	0,06 (0,000)**	0,067 (0,000)**
Haustarifvertrag	0,068 (0,037)*	0,054 (0,033)*	0,043 (0,058)
Konstante	2,218 (0,000)**	1,896 (0,000)**	2,073 (0,000)**
Branchendummies	Ja	Ja	Ja
Beobachtungen	646	646	646
Adj R ² /Pseudo R ²	0,649	0,67	0,45

Quelle: Verdienststrukturerhebung Baden-Württemberg 2001, IAW-Berechnungen, P-Werte in Klammern, * signifikant auf dem 5 % Niveau, ** signifikant auf dem 1 % Niveau.

2.2 Determinanten der betrieblichen Lohnhöhe, Ergebnisse von OLS-Regressionen, Robusten Regressionen und Medianregressionen für Angestellte			
Merkmal	Angestellte		
	OLS	Robuste Regression	Medianregression
Alter/10	0,274 (0,002)**	0,264 (0,000)**	0,328 (0,001)**
(Alter/10) ²	-0,03 (0,004)**	-0,029 (0,000)**	-0,037 (0,002)**
Betriebszugehörigkeitsdauer/10	0,034 (0,296)	0,033 (0,000)**	0,033 (0,013)*
(Betriebszugehörigkeitsdauer/10) ²	-0,002 (0,851)	-	-
Jahre schul./berufl. Ausbildung	0,025 (0,000)**	0,024 (0,000)**	0,025 (0,000)**
Frauenanteil	-0,196 (0,000)**	-0,194 (0,000)**	-0,197 (0,000)**
Ang. mit einf. Tätigkeit (Referenzkategorie)	-	-	-
Anteil Ang. mit besonderen Fachkenntnissen	0,61 (0,000)**	0,427 (0,000)**	0,427 (0,000)**
Anteil Ang. mit selbstständiger Tätigkeit	0,642 (0,000)**	0,447 (0,000)**	0,461 (0,000)**
Anteil Ang. ohne Entscheidungsbefugnis	0,365 (0,003)**	0,188 (0,001)**	0,197 (0,031)*
Entlohnung für technische Tätigkeit	0,008 (0,653)	0,028 (0,066)	0,027 (0,287)
Sonntagsarbeit	0,06 (0,094)*	0,064 (0,11)	0,096 (0,14)
Nachtarbeit	0,054 (0,166)	0,031 (0,417)	0,012 (0,844)
Betriebsgröße 200 – 499	0,01 (0,325)	0,013 (0,147)	0,013 (0,353)
Betriebsgröße 500 – 999	0,018 (0,238)	0,026 (0,021)*	0,02 (0,275)
Betriebsgröße 1 000 und mehr	0,032 (0,015)*	0,036 (0,001)**	0,035 (0,061)
Flächentarifvertrag	0,129 (0,000)**	0,14 (0,000)**	0,144 (0,000)**
Haustarifvertrag	0,132 (0,000)**	0,132 (0,000)**	0,147 (0,000)**
Konstante	1,429 (0,000)**	1,634 (0,000)**	1,494 (0,000)**
Branchendummies	ja	ja	ja
Beobachtungen	667	667	667
Adj R ² /Pseudo R ²	0,744	0,794	0,543

Quelle: Verdienststrukturerhebung Baden-Württemberg 2001, IAW-Berechnungen, P-Werte in Klammern, * signifikant auf dem 5 % Niveau, ** signifikant auf dem 1 % Niveau.

Der Frauenanteil ist ebenfalls sowohl bei den Arbeitern als auch den Angestellten ein statistisch hochsignifikanter Bestimmungsfaktor von betrieblichen Lohnunterschieden: Betriebe mit einem überdurchschnittlichen Frauenanteil verzeichnen unter sonst gleichen Bedingungen sowohl bei den Arbeitern als auch den Angestellten ein geringeres Lohnniveau – und das, obwohl die Auswertungen sich auf Vollzeitbeschäftigte beziehen und weitere systematische Einflussfaktoren von Lohnunterschieden kontrolliert werden. Dabei ist eine überdurchschnittliche Frauenquote im Betrieb bei den Angestellten sogar in noch stärkerem Maße als bei Arbeitern für ein niedrigeres Lohnniveau verantwortlich.

Betriebe mit überdurchschnittlich vielen Arbeiter/-innen die nicht nach Zeitlohn, sondern durch leistungsabhängige Vergütungsformen entlohnt werden, weisen ebenso unter sonst gleichen Bedingungen ein höheres Lohnniveau auf wie Betriebe, in denen Zulagen für belastende Arbeitsbedingungen gezahlt werden. Zulagen für belastende Arbeitsbedingungen sind bei Angestellten dagegen nur wenig verbreitet. Auch lässt sich hier kein signifikanter Zusammenhang konstatieren.

Die Betriebsgröße spielt insbesondere bei der Entlohnung der Arbeiter/-innen, weniger bei der Entlohnung der Angestellten eine Rolle. In Großunternehmen mit mindestens 1 000 Beschäftigten ist das betriebliche Lohnniveau bei den Arbeiterinnen und Arbeitern c. p. um rund 7 bis 8 % höher als in Betrieben mit 100 bis 199 Beschäftigten.

Die Branchendummyvariablen auf Zweistellerebene der Wirtschaftszweigsystematik WZ 93 sind gemeinsam höchstsignifikant, sodass erwartungsgemäß Brancheneffekte in entscheidendem Maße die betriebliche Lohnhöhe bei der Entlohnung der Arbeiter/-innen bestimmen. Auf die Darstellung der Schätzergebnisse für die Branchenvariablen wurde aus Gründen der Übersichtlichkeit verzichtet. Deutlich wird, dass die Referenzbranche Fahrzeugbau üblicherweise in Baden-Württemberg sowohl bei den Arbeitern als auch den Angestellten zu den Branchen mit dem höchsten Durchschnittslohniveau zählt. Das Ledergewerbe, die Textil- und Bekleidungsindustrie sowie das Ernährungsgewerbe sind bei Arbeiterinnen, Arbeitern und Angestellten jeweils am Ende des Feldes zu finden.

Zusammenfassend kann festgehalten werden, dass tarifgebundene Betriebe unter ansonsten gleichen Bedingungen sowohl für Arbeiter/-innen als auch für Angestellte im Produzierenden Gewerbe Baden-Württembergs höhere Löhne zahlen. Insofern deutet das Ergebnis an, dass Tarifbindung zu einem höheren betrieblichen Lohnniveau führt. Kritisch anzumerken ist jedoch, dass letztlich die Richtung der Kausalität aufgrund von Endogenitätsproblemen nicht abschließend geklärt werden kann, was jedoch in den abschließenden Bemerkungen noch thematisiert wird.

4.3 Tarifbindung und Lohnstreuung

Die Kerndichteschätzungen der Löhne in Abschnitt 4.1 haben gezeigt, dass die Streuung der Löhne sowohl bei den Arbeiterinnen und Arbeitern als auch den Angestellten in Betrieben mit individuellen Vereinbarungen am größten ist. Ein solches Ergebnis könnte damit erklärt werden, dass bei Nicht-Orientierung an den Vorgaben eines Tarifvertrags gerade nach unten größere

Abweichungen zulässig sind und somit ergänzendes Differenzierungspotenzial besteht. Andererseits ist zu bedenken, dass in einigen Tarifverträgen bereits Öffnungsmöglichkeiten bestehen, die ein höheres Maß an Lohnflexibilität zulassen, worüber jedoch leider in der Verdienststrukturerhebung nichts bekannt ist. Darüber hinaus orientieren sich Betriebe mit individuellen Lohnvereinbarungen häufig an Flächentarifverträgen (vgl. Strotmann/Vogel 2004), worüber in der Verdienststruktur ebenfalls keine Informationen verfügbar sind.

Die Vorgehensweise bei der Untersuchung möglicher Determinanten der betrieblichen Lohnstreuung entspricht dem Vorgehen bei der obigen Analyse der Determinanten der betrieblichen Lohnhöhe. Die Standardabweichung der logarithmierten Bruttostundenlöhne innerhalb des i -ten Betriebs $\sigma_{i,\ln w}$ wird dabei auf einen Vektor erklärender Variablen z_i regressiert, der die Standardabweichungen der oben angeführten stetigen sowie die Mittelwerte der dichotomen Kontrollvariablen enthält. γ ist der zu schätzende Koeffizientenvektor und η_i der entsprechende Störterm:

$$(2) \quad \sigma_{i,\ln w} = f(z_i' \gamma) + \eta_i$$

Wiederum werden die Analysen getrennt für Arbeiter/-innen und Angestellte durchgeführt, wobei als Schätzverfahren neben OLS-Regressionen ergänzend erneut auch robuste Regressionen sowie Medianregressionen eingesetzt werden, um die Stabilität der Schlussfolgerungen gegenüber dem Vorliegen von Ausreißern abzusichern.

Allgemein lässt sich zunächst anführen, dass sowohl für Arbeiter/-innen als auch für Angestellte die Modellanpassung erheblich schlechter ist als bei den Modellen zur Erklärung der betrieblichen Lohnhöhe. Es ist somit deutlich schwieriger, die betriebliche Lohnstreuung zu erklären als die betriebliche Lohnhöhe. Auch sind die Ergebnisse zwischen den Arbeiter/-innen und den Angestellten keineswegs so ähnlich, wie dies noch bei der Analyse der Lohnhöhe der Fall war.

Mit Blick auf die zentrale Fragestellung des möglichen Einflusses der Tarifbindung auf die betriebliche Lohnstreuung ist interessant, dass sich ein statistisch signifikanter Einfluss nur für die Arbeiter/-innen, nicht jedoch für die Angestellten nachweisen lässt. Während somit die betriebliche Lohnstreuung der Arbeiter in Betrieben mit Branchentarif unter sonst gleichen Bedingungen geringer ausfällt als in Betrieben mit individuellen Lohnvereinbarungen, lässt sich eine entsprechende Lohnkompression durch Flächentarifverträge bei den Angestellten nicht beobachten.

Übereinstimmend gelangen die Schätzungen für beide Gruppen zu dem Ergebnis, dass eine überdurchschnittliche Streuung des Alters in einem Betrieb mit einer höheren Streuung der Bruttostundenlöhne einhergeht. Ein Einfluss der Betriebsgröße auf die Lohnstreuung kann dagegen in keinem Fall nachgewiesen werden. Brancheneinflüsse sind wie bei der Erklärung der betrieblichen Lohnhöhe erneut gemeinsam signifikant, wobei die Lohnstreuung bei den Arbeiter/-innen und Angestellten c. p. im Energiegewerbe, der Papierindustrie und in der Chemieindustrie besonders hoch ausfällt.

3.1 Determinanten der betrieblichen Lohnstreuung, Ergebnisse von OLS-Regressionen, Robusten Regressionen und Medianregressionen für Arbeiter/-innen			
Merkmal	Arbeiter/-innen		
	OLS	Robuste Regression	Medianregression
Alter/10	0,094 (0,027)*	0,086 (0,008)**	0,087 (0,088)
(Alter/10) ²	-0,006 (0,219)	-0,007 (0,055)	-0,005 (0,326)
Betriebszugehörigkeitsdauer/10	0,008 (0,458)	0,015 (0,030)*	0,023 (0,038)*
Verheiratete	0,029 (0,072)	0,012 (0,366)	0,02 (0,329)
Jahre schul./berufl. Ausbildung	0,002 (0,78)	0,003 (0,568)	-0,001 (0,943)
Frauenanteil	0,046 (0,003)**	0,055 (0,000)**	0,05 (0,002)**
Anteil qualifizierter Facharbeiter/-innen	0,035 (0,008)**	0,035 (0,001)**	0,03 (0,073)
Anteil Facharbeiter/-innen	0,005 (0,717)	0,007 (0,493)	0,009 (0,578)
Anteil angelernter Arbeiter/-innen	-0,015 (0,188)	-0,014 (0,149)	-0,013 (0,409)
Prämienlohn	-0,003 (0,701)	-0,006 (0,506)	-0,008 (0,569)
Akkordlohn	-0,012 (0,355)	-0,014 (0,197)	-0,013 (0,43)
Prämien und Akkordlohn	0,003 (0,924)	-0,000 (0,998)	0,056 (0,223)
Mischlohn	-0,008 (0,444)	-0,008 (0,466)	-0,013 (0,445)
Sonntagsarbeit	0,009 (0,366)	0,007 (0,411)	0,007 (0,58)
Nachtarbeit	-0,007 (0,305)	0,000 (0,946)	-0,003 (0,79)
Betriebsgröße 200 – 499	0,002 (0,712)	-0,002 (0,605)	-0,005 (0,528)
Betriebsgröße 500 – 999	0,000 (0,95)	-0,007 (0,272)	-0,005 (0,579)
Betriebsgröße 1 000 und mehr	0,004 (0,531)	0,002 (0,73)	0,003 (0,767)
Flächentarifvertrag	-0,028 (0,000)**	-0,029 (0,000)**	-0,031 (0,000)**
Haustarifvertrag	-0,026 (0,047)*	-0,027 (0,012)*	-0,019 (0,241)
Konstante	0,086 (0,000)**	0,103 (0,000)**	0,087 (0,002)**
Branchendummies	ja	ja	ja
Beobachtungen	642	642	642
Adj R ² /Pseudo R ²	0,144	0,173	0,116

Quelle: Verdienststrukturerhebung Baden-Württemberg 2001, IAW-Berechnungen, P-Werte in Klammern, * signifikant auf dem 5 % Niveau, ** signifikant auf dem 1 % Niveau.

3.2 Determinanten der betrieblichen Lohnstreuung, Ergebnisse von OLS-Regressionen, Robusten Regressionen und Medianregressionen für Angestellte

Merkmal	Angestellte		
	OLS	Robuste Regression	Medianregression
Alter/10	0,186 (0,000)**	0,244 (0,000)**	0,239 (0,000)**
(Alter/10) ²	-0,014 (0,005)**	-0,021 (0,000)**	-0,02 (0,002)**
Betriebszugehörigkeitsdauer/10	0,003 (0,92)	0,004 (0,607)	0,006 (0,639)
(Betriebszugehörigkeitsdauer/10) ²	-0,001 (0,85)	-	-
Jahre schul./berufl. Ausbildung	0,006 (0,047)*	0,007 (0,002)**	0,008 (0,046)*
Frauenanteil	0,016 (0,347)	0,015 (0,279)	0,024 (0,306)
Anteil Ang. mit bsd. Fachkenntnissen	-0,195 (0,004)**	-0,278 (0,000)**	-0,245 (0,000)**
Anteil Ang. mit selbstständiger Tätigkeit	-0,232 (0,001)**	-0,307 (0,000)**	-0,276 (0,000)**
Anteil Ang. ohne Entscheidungsbefugnis	-0,17 (0,015)*	-0,256 (0,000)**	-0,223 (0,001)**
Entlohnung für technische Tätigkeit	0,012 (0,366)	0,005 (0,658)	0,005 (0,782)
Sonntagsarbeit	0,025 (0,416)	0,02 (0,482)	0,03 (0,49)
Nachtarbeit	-0,042 (0,144)	-0,034 (0,228)	0,048 (0,293)
Betriebsgröße 200 – 499	-0,002 (0,77)	-0,008 (0,229)	-0,009 (0,386)
Betriebsgröße 500 – 999	0,001 (0,89)	-0,014 (0,088)	-0,012 (0,332)
Betriebsgröße 1 000 und mehr	-0,005 (0,5)	-0,013 (0,113)	-0,012 (0,332)
Flächentarifvertrag	0,002 (0,759)	0,002 (0,815)	-0,006 (0,586)
Haustarifvertrag	0,02 (0,218)	0,019 (0,184)	0,005 (0,819)
Konstante	0,311 (0,000)**	0,385 (0,000)**	0,363 (0,000)**
Branchendummies	ja	ja	ja
Beobachtungen	655	655	655
Adj R ² /Pseudo R ²	0,145	0,235	0,139

Quelle: Verdienststrukturerhebung Baden-Württemberg 2001, IAW-Berechnungen, P-Werte in Klammern, * signifikant auf dem 5 % Niveau, ** signifikant auf dem 1 % Niveau.

5 Zusammenfassung und Ausblick

Ziel des vorliegenden Beitrags ist es, anhand von Mikrodaten aus der Verdienststrukturerhebung 2001 für das Produzierende Gewerbe Baden-Württembergs die Frage zu beantworten, welchen Einfluss kollektive Lohnfindungsregelungen auf die betriebliche Lohnhöhe und die betriebliche Lohnstreuung haben. Auf der Grundlage verschiedener multivariater Analysemethoden wird dabei gezeigt, dass ansonsten vergleichbare Betriebe, die der Tarifbindung unterliegen, *ceteris paribus* höhere Löhne zahlen als Betriebe mit individuellen Lohnvereinbarungen. Der Effekt der Tarifbindung ist dabei bei den Angestellten größer als bei den Arbeiter/-innen. Eine die Lohnstreuung reduzierende Wirkung der Tarifbindung kann für Arbeiter/-innen, jedoch nicht für die Angestellten nachgewiesen werden.

Abschließend soll auf einige methodische Probleme hingewiesen sowie ein Ausblick auf weiteren Forschungsbedarf gegeben werden. Ein besonderes Problem der bloßen Verwendung von Querschnittsdaten bei der Analyse der Bedeutung der Tarifbindung für die Lohnhöhe und die Lohnstreuung ist die Endogenität der Tarifbindung als erklärende Variable. Während die obigen Modelle unterstellen, dass die Kausalität von der Tarifbindung zur Lohnhöhe geht, ist inhaltlich auch eine umgekehrte Kausalität denkbar: So ist vorstellbar, dass sich gerade die Betriebe, in denen ein höheres Lohnniveau gezahlt wird, weiterhin der Tarifbindung unterziehen, während Betriebe, die sich nur ein geringeres Lohnniveau leisten können, nicht an kollektive Regelungen binden. Um diese Endogenitätsproblematik lösen zu können, sind Paneldaten erforderlich, eine geeignete Instrumentierung ist allein mit dem Querschnitt 2001 nicht sinnvoll möglich. Wünschenswert wäre nicht zuletzt daher, verschiedene Wellen der Verdienststrukturerhebung zukünftig zu einem Betriebspanel zu verknüpfen, um somit auch Entwicklungen im Zeitablauf verfolgen und analysieren zu können.

Von besonderem Interesse wäre es auch, die Bedeutung des Motivs der Effizienzlöhne, nach dem höhere Löhne aus betrieblichem Eigeninteresse gezahlt werden, um die Motivation und damit die Produktivität der Mitarbeiter/-innen zu steigern, isolieren zu können (vgl. dazu z. B. Franz/Pfeiffer 2003).

Entsprechende Analysen zu den Zusammenhängen zwischen Tarifbindung und Lohnhöhe und Lohnstreuung auf der Individualebene anstelle der betrieblichen Ebene wurden ergänzend von Bechtel (2004) durchgeführt, wobei die inhaltlichen Ergebnisse weitgehend identisch ausfielen. Die Zusammenhänge könnten dabei auch an unterschiedlichen Enden der Lohnverteilung unterschiedlich ausfallen, so dass in weiteren Analysen nicht nur Medianregressionen, sondern auch allgemeinere Quantilsregressionen für andere Quantile der Verteilung durchgeführt werden sollten.¹³⁾

Ein zentraler ergänzender Forschungsbedarf ist die Bedeutung von Öffnungs- und Härteklauseln in Tarifverträgen und deren Einfluss auf die Lohnhöhe und die Lohnstreuung. So ist unter anderem von Interesse, ob in Betrieben, deren Tarifverträge Öffnungsmöglichkeiten vorsehen, gegenüber tarifgebundenen Betrieben, deren Tarifvereinbarungen nicht über entsprechende Öffnungsmöglichkeiten verfügen, ein nachweisbarer Effekt in Richtung einer geringeren Lohnhöhe und zusätzlicher Lohnstreuung zu beobachten ist. Bezüglich der Lohnhöhe ist der Effekt

13) Vgl. dazu z. B. auch Bechtel (2004).

aus theoretischer Sicht nicht eindeutig, da die Einführung von tariflichen Öffnungsklauseln durchaus dazu führen kann, dass die Gewerkschaften eine Art Lohndifferenzierung verfolgen. Dabei werden bei ertragsstärkeren Betrieben Verteilungsspielräume abgeschöpft, wohlwissend, dass einige Betriebe durch die Öffnungsmöglichkeit ein geringeres Lohnniveau verhandeln werden (vgl. dazu Fitzenberger/Franz 1999). Das IAW Tübingen widmet sich der Analyse dieses Themas derzeit im Rahmen eines DFG-Forschungsprojekts im DFG-Arbeitsmarktschwerpunkt „Flexibilisierungspotenziale bei heterogenen Arbeitsmärkten“. Auf der Grundlage einer Erhebung der Öffnungs- und Härtefallklauseln in den Tarifbereichen in Baden-Württemberg sollen der Verdienststrukturerhebung auf Betriebsebene Informationen über das Ausmaß der Öffnung der Tarifverträge zugespielt werden.¹⁴⁾ Darüber hinaus ist geplant, die Bedeutung des Ausmaßes der Tariföffnung und die tatsächliche Inanspruchnahme von Öffnungsklauseln durch die Betriebe und deren Wirkungen auf die Lohnstruktur der Betriebe und die betriebliche Entwicklung mit Hilfe des Linked-Employer-Employee-Datensatzes des IAB Nürnberg (vgl. Alda 2005) zu untersuchen.

Literatur

Abowd, J. M./Kramarz, F. (1999): The Analysis of Labor Markets using Matched Employer-Employee Data, in: Ashenfelter, O. C./Card, David (Hrsg.): Handbook of Labor Economics, Vol. 3B, Amsterdam: Elsevier, S. 2629 – 2710.

Acemoglu, D./Aghion, P./Violante, G. L. (2001): Deunionization, Technical Change and Inequality, in Carnegie-Rochester Conference Series of Public Policy 55, S. 229 – 264.

Alda, H. (2005): Betriebe und Beschäftigte in den Linked-Employer-Employee Daten. LIAB des Instituts für Arbeitsmarkt- und Berufsforschung. Forschungsdatenzentrum (FDZ) der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB) Datenreport Nr. 1/2005.

Arndt, C. (2004): Tarifbindung und Tarifflicht – Welche Faktoren sind entscheidend? In: IAW-Report Heft 2/2004, S. 63 – 92.

Artus, I. (2001): Krise des deutschen Tarifsystems. Die Erosion des Flächentarifvertrags in Ost und West, Wiesbaden: Westdeutscher Verlag.

Bechtel, S. (2004): Verdienststrukturen in Baden-Württemberg: Eine empirische Analyse mit Hilfe von Quantilsregressionen, mimeo.

Becker, G. S. (1975): Human Capital: A Theoretical and Empirical Analysis, 2. Aufl., New York: National Bureau of Economic Research.

¹⁴⁾ Erste Ergebnisse zur Einführung verschiedener Öffnungsklauseln im Bereich des Produzierenden Gewerbes in Baden-Württemberg und das Ausmaß der Öffnung in verschiedenen Branchen werden in Heimbach (2005) vorgestellt.

Bispinck, R. (2003): Das deutsche Tarifsysteem in Zeiten der Krise – Streit um Flächentarif, Differenzierung und Mindeststandards, in: WSI-Mitteilungen, Heft 7/2003, S. 395 – 404.

Blau, F. D./Kahn, L. M. (1999): Institutions and Laws in the Labor Market. In: Ashenfelter, O./Card, D. (Hrsg.): Handbook of Labor Economics, Vol. 3A, Amsterdam: Elsevier, 1399 – 1461.

Brown, C./Hamilton, J./Medoff, J. (1990): Employers Large and Small, Cambridge: Harvard University Press.

Fitzenberger, B./Franz, W. (1999): Der Flächentarifvertrag: Eine kritische Würdigung aus ökonomischer Sicht, in: Franz, W./Hesse, H./Ramser, H.J./Stadler, M. (Hrsg.): Ökonomische Analyse von Verträgen, Schriftenreihe des wirtschaftswissenschaftlichen Seminars Ottobeuren. Tübingen: J.C.B. Mohr (Paul Siebeck), 2000, S. 191 – 232.

Franz, W. (2003): Arbeitsmarktökonomik, 5. Auflage, Heidelberg u.a. : Springer-Verlag.

Franz, W. (1995): Die Lohnfindung in Deutschland in einer internationalen Perspektive: Ist das deutsche System ein Auslaufmodell?, in: Beihefte der Konjunkturpolitik, Heft 43, S. 31 – 57.

Franz, W./Pfeiffer, F. (2003): Tarifbindung und die ökonomische Rationalität von Lohnrigiditäten, in Jahrbücher für Nationalökonomie und Statistik, 233(1), S.23 – 57.

Freeman, R. B. (1982): Union Wage Practices and Wage Dispersion within Establishments. Industrial and Labor Relations Review, 36, S. 3 – 21.

Freeman R. B./Medoff J. L. (1984): What Do Union Do? New York: Basic Books.

Gerlach, K./Stephan, G. (2002): Tarifverträge und Lohnstruktur in Niedersachsen. Ein Blick zurück: Die Gehalts- und Lohnstrukturerhebungen 1990 und 1995, in: Statistische Monatshefte Niedersachsen, Heft 10/2002, S. 543 – 552.

Gerlach, K./Stephan, G. (2005): Wage Distributions by Wage-Setting Regime. IAB-Diskussionspapier Nr. 9/2005, Nürnberg.

Hafner, H. P. (2006): Erhebungen zu Verdiensten und Arbeitskosten. Beitrag in diesem Band

Heinbach, W. D. (2005): Ausmaß und Grad der tarifvertraglichen Öffnung; in IAW-Report Heft 2/2005, S. 51 – 70.

Kohaut, S./Schnabel, C. (2003): Tarifverträge – Nein danke!? Ausmaß und Einflussfaktoren der Tarifbindung west- und ostdeutscher Betriebe, in: Jahrbücher für Nationalökonomie und Statistik, 223(3), S. 312 – 331.

Lemieux, T. (1998): Estimating the Effects of Unions on Wage Inequality in a Panel Data Model with Comparative Advantage and Nonrandom Selection, Journal of Labor Economics 16, S. 261 – 291.

Lindbeck, A./Snower, D. J. (2001): Centralized Bargaining and Reorganized Work: Are they Compatible?, *European Economic Review* 45, S. 1885 – 1875.

Mincer, J. (1974): *Schooling, Experience, and Earnings*, New York: National Bureau of Economic Research

Mödinger, P. (2003): Verdienststrukturerhebung 2001, in: *Statistisches Monatsheft Baden-Württemberg*, Heft 10/2003, S. 3 – 11.

Rousseeuw, P.J./Leroy, A.M. (1987): *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.

Statistisches Landesamt Baden-Württemberg (Hrsg.) (2004): *Tarif- und Lohnstrukturen in Baden-Württemberg*. Statistische Analysen 7/2004.

Strotmann, H./Vogel, A. (2004): Entwicklung der Tarifbindung baden-württembergischer Betriebe in den Jahren 2000 bis 2003. *IAW-Kurzbericht* 3/2004.

Strotmann, H. (2005): Determinanten der betrieblichen Einführung von Gewinnbeteiligungsmo-
dellen – eine empirische Analyse mit Betriebspaneldaten, in: *Zeitschrift für Betriebswirtschaft*,
75, Heft 12, S. 1 – 29.

White, H. (1980): A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct
Test for Heteroscedasticity, in: *Econometrica*, 48(4), S. 817 – 838.

Autoren- und Kontaktangaben:

Dr. Harald Strotmann, Institut für Angewandte Wirtschaftsforschung (IAW) Tübingen, Geschäfts-
führung, Ob dem Himmelreich 1, 72074 Tübingen, E-Mail: harald.strotmann@iaw.edu

Diplom-Ökonom Stephan Bechtel, TNS Infratest GmbH, Landsberger Str. 338, D-80687 Mün-
chen, E-Mail: stephan.bechtel@tns-infratest.com

Diplom-Ökonom Wolf Dieter Heinbach, Wissenschaftlicher Referent am IAW Tübingen, Ob dem
Himmelreich 1, 72074 Tübingen, E-Mail: wolf.heinbach@iaw.edu

Erhebungen zu Verdiensten und Arbeitskosten

Hans-Peter Hafner

Der regionale Standort Wiesbaden ist im Rahmen des Forschungsdatenzentrums der Statistischen Landesämter fachlich für den Bereich „Verdienste und Arbeitskosten“ zuständig. Seit Ende 2004 steht auf der Datenbasis der Gehalts- und Lohnstrukturerhebung 2001 Wissenschaftlern ein Linked-Employer-Employee-Datensatz zur Verfügung, der Analysen zu Verdienstdifferenzen unter Einbeziehung betriebsspezifischer Charakteristika ermöglicht. Seit September 2005 sind die Daten der Arbeitskostenerhebung 2000 nutzbar. Sie ermöglichen u. a. eine Untersuchung der Lohnnebenkosten, differenziert nach Wirtschaftszweigen. In diesem Aufsatz geben wir zunächst einen Überblick über Methodik und Merkmale der Gehalts- und Lohnstrukturerhebung, stellen anschließend einige der bisherigen Forschungsarbeiten vor, präsentieren ausgewählte Ergebnisse der Erhebung 2001 und beschreiben die unterschiedlichen Möglichkeiten des Datenzugangs für Wissenschaftler. Im zweiten Teil gehen wir dann auf Methodik und Merkmale der Arbeitskostenerhebung ein.

1 Die Gehalts- und Lohnstrukturerhebung

1.1 Methodik und Merkmale

Die Gehalts- und Lohnstrukturerhebung (GLS) wird von den statistischen Ämtern des Bundes und der Länder seit 1951 durchgeführt. Nachdem die Erhebung in der Vergangenheit in unregelmäßigen Abständen mit einer großen Lücke zwischen 1978 und 1990 stattfand, ist die Periodizität zukünftig vierjährlich. Auf Grundlage einer Verordnung der EG von 1999 findet die Erhebung in allen EU-Ländern statt. Somit liegen europaweit vergleichbare Daten vor. Da die meisten Länder die letzte Erhebung für das Jahr 2002 durchführten, findet die nächste somit für 2006 statt. Die Angaben zu Arbeitszeit und Verdienst beziehen sich immer auf den Berichtsmonat Oktober.

Zum Berichtskreis gehören Betriebe des Produzierenden Gewerbes und ausgewählte Teile des Dienstleistungsbereichs. Neben dem Handel und dem Kredit- und Versicherungsgewerbe gibt es seit 2001 Daten für das Gastgewerbe, Verkehr und Nachrichtenübermittlung, das Grundstücks- und Wohnungswesen, die Vermietung beweglicher Sachen und die Erbringung von Dienstleistungen überwiegend für Unternehmen. Auch bei den Beschäftigten erfolgte eine Ausweitung gegenüber früheren Erhebungen: Auszubildende, geringfügig Beschäftigte und Arbeitnehmer in Altersteilzeit sind erstmals 2001 erfasst worden.

Die GLS ist eine zweistufige Stichprobenerhebung. In der ersten Auswahlstufe wird aus den Betrieben eine geschichtete Zufallsstichprobe gezogen. Die Schichtung erfolgt dabei nach 17 Regionen (Bundesländer, Berlin getrennt nach West und Ost), 64 Wirtschaftsgruppen und 7 Beschäftigtengrößenklassen. In der zweiten Stufe werden in den ausgewählten Betrieben die einzubeziehenden Beschäftigten über die Personalnummer aus den Personallisten ermittelt. Dazu

werden von den statistischen Ämtern eine Startzahl und ein Auswahlabstand vorgegeben. Insgesamt lieferten für 2001 gut 22 000 Betriebe Angaben zu über 845 000 Beschäftigten. Die Verteilung des Stichprobenumfangs auf die Bundesländer erfolgt so, dass die Ergebnisse für jede Region einen vergleichbaren Standardfehler besitzen. D. h.: Der Auswahlsatz ist für die kleinen Bundesländer höher als für die großen. Für eine tiefere regionale Untergliederung als nach Bundesländern sind keine repräsentativen Aussagen möglich. Da im Datensatz aber auch die Gemeindeganzzahl des Betriebssitzes enthalten ist, können darüber z. B. zusätzliche Informationen auf Kreisebene angespielt werden wie etwa die Siedlungsstrukturtypen des Bundesamtes für Bauwesen und Raumordnung (BBR) (siehe www.bbr.bund.de – Themen Projekte – Raumbeobachtung) oder regionale Arbeitslosenquoten.

Es gibt getrennte Erhebungsbögen für die Angaben zum Betrieb und jeweils einen (oder bei größeren Betrieben mehrere) Bögen für die Arbeiter und die Angestellten. Für die Betriebe bzw. das Unternehmen, zu dem der Betrieb gehört, liegen die folgenden Angaben vor:

- Wirtschaftszweig nach WZ93 (deutsche Systematik, von NACE Rev. 1 abgeleitet)¹⁾
- Einfluss der öffentlichen Hand auf die Unternehmensführung: Kein Einfluss, eingeschränkter Einfluss bis 50 %, beherrschender Einfluss über 50 %
- Anzahl Beschäftigte des Unternehmens
- Anzahl Beschäftigte des Betriebes differenziert nach Arbeitern und Angestellten und Geschlecht
- Schlüssel der Tarifverträge, die für den Betrieb gelten: Daraus kann man u. a. ablesen, ob für den Betrieb ein Kollektiv- oder ein Firmentarifvertrag oder eine Betriebsvereinbarung gilt
- Hochrechnungsfaktoren: Diese Faktoren dienen zur freien Hochrechnung der Daten auf die Grundgesamtheit der Betriebe bzw. der Beschäftigten. Sie ergeben sich für die Betriebe als Quotient der Anzahl der Unternehmen in der Schicht in der Grundgesamtheit und der Anzahl der verwertbaren Antworten in der Schicht, für die Arbeitnehmer als Quotient aus der Anzahl der Beschäftigten des Betriebes insgesamt und der Anzahl der in die Erhebung einbezogenen Beschäftigten.

Für alle Beschäftigten werden die folgenden Angaben erfasst:

- Tarifliche Gehalts- oder Lohngruppe, falls Verdienst nach Tarifvertrag bezahlt wird
- Leistungsgruppe: Bei Bezahlung nach Tarifvertrag werden die einzelnen Gehaltsklassen den Leistungsgruppen zugeordnet. Erfolgt die Bezahlung nach freier Vereinbarung, so muss der Betrieb anhand der ausgeübten Tätigkeit entscheiden, zu welcher Leistungsgruppe der Beschäftigte gehört. Bei den Angestellten erfolgt eine Differenzierung nach 6 Gruppen, bei den Arbeitern nach 4.
- Soziodemografische Merkmale: Geschlecht, Geburtsmonat und -jahr
- Angaben zu Tätigkeit und Qualifikation: Der fünfstellige Schlüssel aus dem Versicherungsnachweis der Sozialversicherung wird in den Bogen eingetragen. Die ersten drei Stellen bezeichnen die ausgeübte Tätigkeit, die vierte die Stellung im Beruf und die letzte die Ausbildung.
- Arbeitszeit: Vertraglich vereinbarte wöchentliche Arbeitszeit, Mehrarbeitsstunden
- Verdienst: Bruttomonatsverdienst, Verdienst für Mehrarbeit, Zulagen für Schicht-, Nacht- und Sonntags-/Feiertagsarbeit, Bruttojahresverdienst, jährliche Sonderzahlungen
- Urlaubsanspruch in Tagen

1) http://www.forschungsdatenzentrum.de/bestand/gls/2001/fdz_gls2001_11-Klassifikation-WZ93.pdf

Einige zusätzliche Merkmale werden auf Grundlage eines nationalen Gesetzes nur für das Produzierende Gewerbe, Handel und Kredit- und Versicherungsgewerbe erhoben. Dazu gehören Lohnsteuer und Sozialabgaben, Nettoverdienst sowie Steuerklasse und die Anzahl der Kinderfreibeträge.

Da standardmäßig keine Tabellen mit Ergebnissen zu den Kinderfreibeträgen veröffentlicht werden, wird dieses Merkmal nur schlecht gepflegt und im Allgemeinen nicht plausibilisiert.

Weitere Informationen zu Methodik und Merkmalen der GLS 2001 findet man in Frank-Bosch (2003) und in den Metadaten im Internetangebot der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder.²⁾

1.2 Analysen mit der Gehalts- und Lohnstrukturerhebung

Die bisherigen Analysen mit der GLS lassen sich zu einem großen Teil den beiden Schwerpunkten „Geschlechtsspezifische Verdienstunterschiede“ und „Verdiensthöhe in Abhängigkeit von Tarifbindung“ zuordnen. Sie entstanden einerseits im Rahmen eines Kooperationsprojektes der Universität Hannover, des Niedersächsischen Landesamtes für Statistik und des Niedersächsischen Ministeriums für Wirtschaft, Technologie und Verkehr, zum anderen innerhalb eines Pilotprojektes des Statistischen Bundesamtes zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und amtlicher Statistik. Schließlich gab es 2004 eine gemeinschaftliche Analyse des Statistischen Landesamtes Baden-Württemberg mit dem Institut für angewandte Wirtschaftsforschung Tübingen. Erste Projekte im Rahmen des Forschungsdatenzentrums der Statistischen Landesämter laufen derzeit.

1.2.1 Geschlechtsspezifische Verdienstunterschiede

Stephan (1997) untersucht mit den niedersächsischen Daten der GLS 1990, inwieweit die Lohnunterschiede zwischen Männern und Frauen darauf zurückzuführen sind, dass einerseits die Geschlechter einen unterschiedlichen Zugang zu Leistungsgruppen, Betrieben und Arbeitsplatzbereichen haben und dass andererseits das Einkommen bei steigendem Frauenanteil in einem Betrieb oder in einer Leistungsgruppe zurückgeht. Sie verwendet dabei ein von Groshen (1991) entwickeltes Zerlegungsverfahren. Bei diesem wird zunächst mit der Methode der kleinsten Quadrate eine Einkommensfunktion geschätzt, bei der neben einer Dummyvariablen für das Geschlecht die Frauenanteile in der Leistungsgruppe des Beschäftigten, im Betrieb und im Arbeitsplatzbereich als erklärende Variablen aufgenommen werden. Die Differenz der geschätzten Durchschnittseinkommen von Männern und Frauen wird dann zerlegt in einen Teil, der nur vom Geschlecht abhängig ist und drei Komponenten, die von den Frauenanteilen in den Bereichen abhängen.

Es zeigt sich, dass bei steigendem Frauenanteil in den einzelnen Bereichen die Einkommen immer sinken. Da Frauen im Allgemeinen in Bereichen mit einem höheren Frauenanteil arbeiten als Männer, lassen sich dadurch die Lohnunterschiede erklären. Ein gutes Drittel des Lohndiffe-

2) <http://www.forschungsdatenzentren.de/bestand/gls/2001/metadaten.asp>

renzials lässt sich auf den Frauenanteil in der Leistungsgruppe zurückführen, etwa ein Fünftel auf den Anteil im Arbeitsplatzbereich, während der Anteil im Betrieb mit 5 % nur eine geringere Rolle spielt. Das verbleibende gute Drittel kann mit diesem Modell nicht erklärt werden.

Von Kulmiz (2001) untersucht geschlechtsspezifische Lohnunterschiede ebenfalls mit den Daten der GLS 1990, allerdings für das ganze frühere Bundesgebiet. Sie ermittelt mit Hilfe einer für Frauen und Männer getrennt durchgeführten schrittweisen Regression Einkommensfunktionen. Dabei verwendet sie als erklärende Variablen u. a. das Alter, das quadrierte Alter, die Unternehmenszugehörigkeit (linear und quadriert), die Leistungsgruppe, die Ausbildung, 33 Berufsklassen, 25 Wirtschaftszweige, die Arbeitszeit, den aus der Steuerklasse abgeleiteten Familienstand sowie Interaktionsterme aus zum einen Alter und Unternehmenszugehörigkeit und zum anderen Ausbildung und Leistungsgruppe. Die Differenz der durchschnittlichen geschätzten Einkommen für Frauen und Männer wird dann nach dem Verfahren von Oaxaca und Blinder (Oaxaca 1973, Blinder 1973) in einen Ausstattungs- und einen Diskriminierungsanteil zerlegt.

Bei den Arbeitern lassen sich nach der Zerlegung 41 % der Differenz erklären, davon knapp 23 % durch die Leistungsgruppe, 13 % durch den Wirtschaftszweig und 7 % durch die Ausbildung. Für die Angestellten lassen sich sogar 64 % durch die Ausstattung der Beschäftigten bzw. die verwendeten Merkmale des Betriebes erklären. Auch hier geht der größte Teil mit über 35 % auf das Konto der Leistungsgruppe, durch Alter und Wirtschaftszweig lassen sich jeweils 8 % erklären.

Der Einfluss von Tarifverträgen auf Verdienstunterschiede

Stephan und Gerlach (2004) untersuchen mit den Erhebungen 1990, 1995 und 2001 für Niedersachsen Verdienstunterschiede im Verarbeitenden Gewerbe in Unternehmen, die nach Kollektivtarifverträgen, Firmentarifverträgen und nach freier Vereinbarung bezahlen. Sie verwenden ein Mehrebenenmodell, das für die Beschäftigten die Dauer der Unternehmenszugehörigkeit und die aus Ausbildung und Alter errechnete potenzielle Berufserfahrung einbezieht; auf Firmenebene werden außer den Angaben zu den Tarifverträgen der Frauenanteil, die Unternehmensgröße und die durchschnittliche Schulbildung berücksichtigt. Analysiert wird, von welchen Faktoren der Basislohn (Lohn, den der durchschnittliche Beschäftigte im Unternehmen erhält) abhängt. Der Basislohn ist in Firmen, die einen Tarifvertrag anwenden höher, 1990 etwa 3 bis 4 %, 2001 gar 11 bis 12 %. Dabei ist die Differenz bei Verwendung von Kollektivtarifverträgen noch leicht größer als bei Firmentarifverträgen. Von 1990 bis 2001 haben sich aber die durch Qualifikation und Unternehmenszugehörigkeit bedingten Unterschiede verringert. Die Wissenschaftler begründen dies damit, dass bei einer schlechteren wirtschaftlichen Lage, wie sie 2001 im Vergleich zu 1990 bestand, zusätzliche Qualifikationen weniger stark vergütet werden. Weitere Ergebnisse sind, dass der Basislohn bei steigendem Frauenanteil sinkt und bei höherer durchschnittlicher Bildung steigt. Das letztere hängt wahrscheinlich damit zusammen, dass gut zahlende Unternehmen Hochqualifizierte besonders stark anziehen. Der Lohnabstand von Frauen zu Männern ist in Unternehmen mit einem Tarifvertrag ca. 5 % geringer, von 1990 bis 2001 reduzierte er sich in Unternehmen, die nach freier Vereinbarung vergüten, von 21 auf 17 %.

Bechtel, Mödinger und Strotmann (2004) untersuchen eine ähnliche Fragestellung mit den Daten der GLS 1995 und 2001 für Baden-Württemberg. Siehe dazu den Beitrag von Strotmann in diesem Band.

Weitere Forschungsarbeiten

Fitzenberger und Reize (2002) untersuchen für die alten Bundesländer Lohnunterschiede und -wachstum von 1990 bis 1995 differenziert nach verschiedenen Qualifikationsgruppen. Diese Gruppen werden definiert nach den Dimensionen Berufsausbildung (ohne abgeschlossene Berufsausbildung; mit abgeschlossener Berufsausbildung, aber ohne (Fach-) Hochschulabschluss; mit (Fach-) Hochschulabschluss) und Erwerbstyp (Vollzeit erwerbstätige Männer, Vollzeit erwerbstätige Frauen, Teilzeit erwerbstätige Frauen).

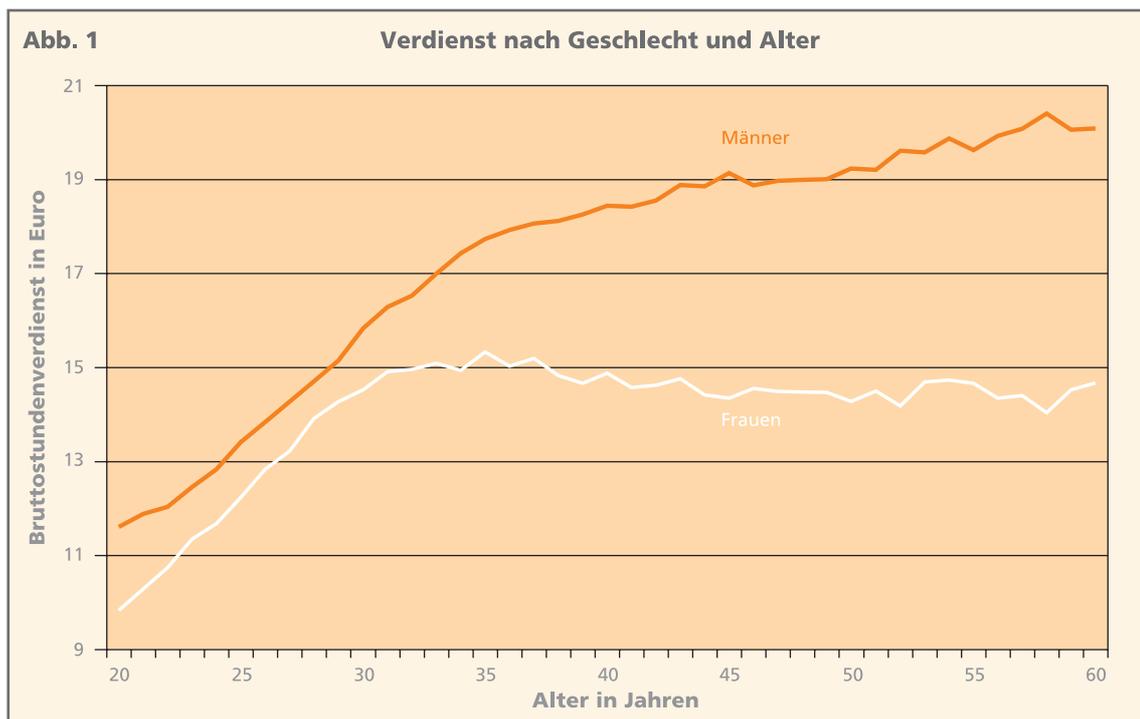
Alle Lohnverteilungen sind eingipflig. Erwartungsgemäß liegt die Verteilung für die Vollzeit erwerbstätigen Männer am weitesten rechts und für die Teilzeit erwerbstätigen Frauen am weitesten links. Die Struktur der Verteilungen ändert sich von 1990 bis 1995 kaum. Die Verdienstunterschiede innerhalb der Gruppen nehmen mit steigendem Qualifikationsniveau zu. Besonders ausgeprägt sind diese Unterschiede bei den Teilzeit erwerbstätigen Frauen, während sie sich zwischen Vollzeit tätigen Männern und Frauen nicht wesentlich unterscheiden.

Jirjahn und Stephan (1999) analysieren mit den Daten für Niedersachsen 1990 und 1995 für männliche Beschäftigte, von welchen Faktoren die Höhe der jährlichen Sonderzahlungen abhängt. Mittels Varianzanalysen wird gezeigt, welche Anteile der Differenzen bei den Sonderzahlungen auf den Beschäftigten, den Betrieb, die Branche und den Arbeitsplatzbereich zurückzuführen sind. Die individuellen Merkmale des Beschäftigten wie Ausbildung und Alter erklären nur um die 10 % der Varianz, während insbesondere die Dauer der Betriebszugehörigkeit eine wichtige Rolle spielt. Durch den Betrieb werden bei den Arbeitern 70 bis 90 %, bei den Angestellten 50 bis 60 % der Varianz erklärt; Sektor und Firmengröße stehen in der Erklärungskraft dagegen weit zurück.

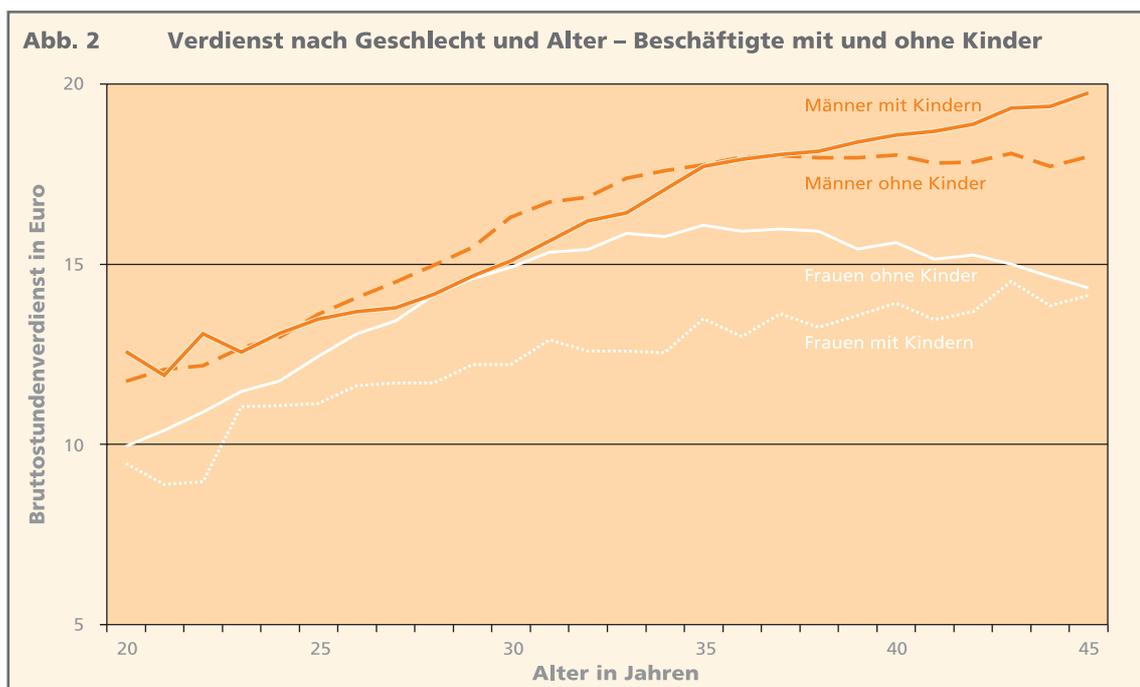
Ausgewählte Ergebnisse der Erhebung 2001

In 2001 verdienen Frauen immer noch weniger als Männer. Um Effekte auszuschließen, die aufgrund schlechterer Bezahlung von Teilzeit entstehen, betrachten wir nur Arbeitsverhältnisse mit einer vertraglich vereinbarten Arbeitszeit von wöchentlich 30 Stunden und mehr. Ferner schließen wir Auszubildende, Beschäftigte in Altersteilzeit und solche mit unbekannter Ausbildung aus den folgenden Betrachtungen aus.

Über alle Alters- und Qualifikationsgruppen hinweg liegt der Bruttostundenlohn von Frauen 20,5 % unter demjenigen der Männer. Dabei liegen die Unterschiede in den ersten Berufsjahren deutlich unter 10 %, z. B. sind es in der Altersgruppe 25 bis 29 gerade 7,3 %. Aber während das Durchschnittsgehalt der Frauen zwischen 30 und 35 Jahren seinen Höhepunkt erreicht und danach sogar leicht zurückgeht, steigt es bei den Männern bis zum Ende des Erwerbslebens.

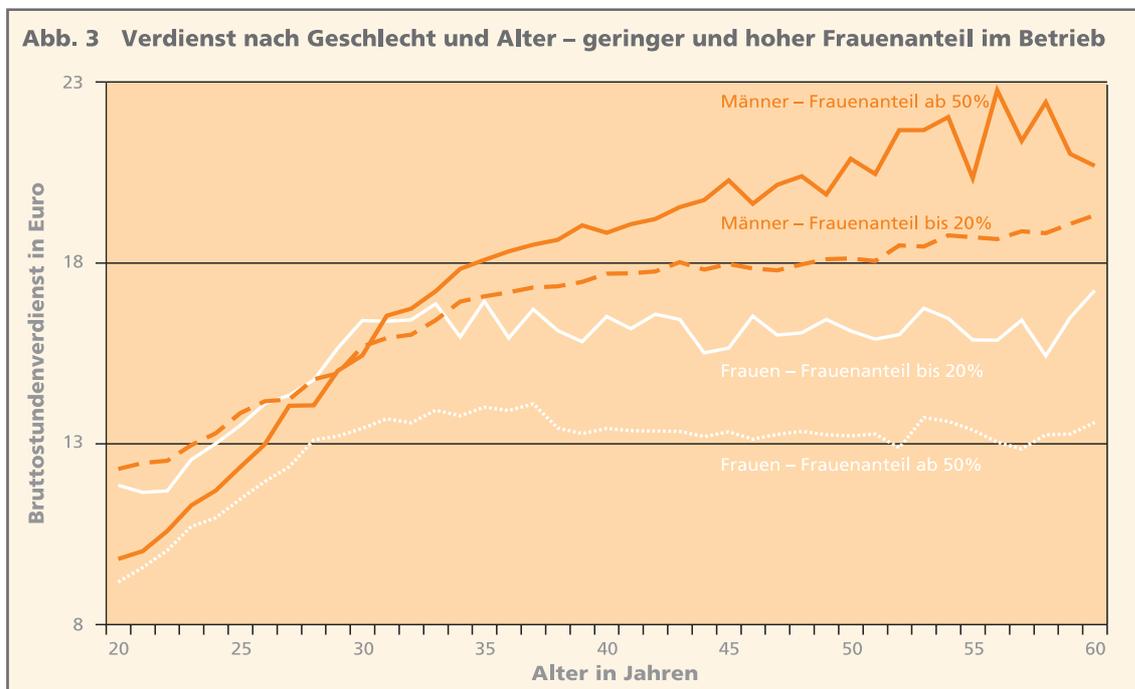


Eine mögliche Erklärung für den größer werdenden Verdienstabstand im Lauf des Erwerbslebens besteht darin, dass aufgrund von Kindererziehung immer noch fast ausschließlich Frauen Unterbrechungen in ihrer Erwerbsbiographie haben. Leider ist die einzige Information zu vorhandenen Kindern die Anzahl der Kinderfreibeträge. Da es bei älteren Frauen anhand der Daten keine Möglichkeit gibt festzustellen, ob sie Kinder haben, betrachten wir nur Frauen bis 45. Bei diesen kann man mit einer großen Wahrscheinlichkeit davon ausgehen, dass die Kinder – soweit vorhanden – noch auf der Steuerkarte eingetragen sind.



Frauen, die Kinder haben, verdienen bereits in der Altersgruppe 25 bis 29 über 17 % weniger als die gleichaltrigen Männer, während der Unterschied bei kinderlosen Frauen gerade 7 % beträgt. Bis zur Gruppe der 40 bis 44jährigen erhöht sich die Differenz beim Vorhandensein von Kindern auf 26,8 %, Frauen ohne Kinder haben in dieser Altersgruppe nur einen Rückstand von 15,5 % auf das Verdienstniveau der Männer.

Auffällig ist ein Unterschied der Gehaltsdifferenzen zwischen Betrieben mit einem niedrigen Anteil weiblicher Beschäftigter (bis 20 %) und solchen, in denen mindestens die Hälfte der Beschäftigten Frauen sind. Bei niedrigem Frauenanteil ist der Abstand zwischen Männern und Frauen relativ gering, in den unteren Altersgruppen verdienen die Frauen teilweise sogar mehr. In den Altersgruppen über 30 verdienen Frauen in Betrieben mit einem hohen Anteil an weiblichen Beschäftigten deutlich weniger als in männerdominierten Betrieben, während es bei Männern genau umgekehrt ist. Für die Erklärung dieses Phänomens sind weitere Analysen erforderlich, z. B. zur Qualifikationsstruktur der weiblichen und männlichen Beschäftigten in den beiden Betriebsgruppen.



Datenzugang für Wissenschaftler

Die Einzeldaten der GLS 2001 für ganz Deutschland stehen seit Ende 2004 zur OnSite-Nutzung an den Gastwissenschaftlerarbeitsplätzen an allen Standorten des FDZ der Statistischen Landesämter und mittels kontrollierter Datenfernverarbeitung zur Verfügung. Die Daten von 1995 und für die alten Bundesländer von 1990 sowie für die neuen Bundesländer von 1992 wurden Ende 2005 aufbereitet. Da bedauerlicherweise einige Materialien von 1990 bereits vernichtet wurden, stehen diese Daten nur für 7 Bundesländer zur Verfügung (nicht mehr vorhanden: Hamburg, Bremen, Hessen, Nordrhein-Westfalen). Für 1992 fehlt Mecklenburg-Vorpommern.

Bei einem Vergleich der aktuellen Daten mit den älteren ist zu berücksichtigen, dass 1990 teils nach der alten Wirtschaftszweigklassifikation für das Verarbeitende Gewerbe (SYPRO)³⁾ und teils nach der Klassifikation WZ79 erhoben wurde; 1995 ist für die Mehrzahl der Betriebe bereits der Code nach WZ93 vorhanden, aber nicht für alle. Für den Umstieg von SYPRO bzw. WZ79 auf WZ93 gibt es zwar Umschlüsselungsbrücken, aber die Zuordnung von alten zu neuen Wirtschaftszweigen ist in vielen Fällen nicht eindeutig möglich. Referenzen zur Umschlüsselung werden demnächst bei den Metadaten zur GLS im Internetangebot des Forschungsdatenzentrums vorhanden sein.

Derzeit läuft ein gemeinsames Projekt des Standortes Wiesbaden mit dem Statistischen Bundesamt zur Erstellung eines Scientific-Use-Files der Erhebung 2001. Es fanden zwei Treffen mit Wissenschaftlern statt, bei denen diese ihre Wünsche im Hinblick auf für geplante Analysen wichtige Merkmale äußern konnten. Momentan werden verschiedene Anonymisierungsvarianten anhand von Angriffsszenarien abschließend auf den Schutz der Daten getestet. Im Scientific-Use-File werden voraussichtlich 5 Regionen ausgewiesen, die aus der Zusammenfassung benachbarter Bundesländer entstehen sowie 30 bis 40 Wirtschaftsgruppen. Weitere Schutzmaßnahmen bestehen u. a. in der Mikroaggregation der Beschäftigtenanzahl für große Unternehmen bzw. Betriebe. Die Veröffentlichung wird voraussichtlich Ende Frühjahr/Anfang Sommer 2006 erfolgen.

2 Die Arbeitskostenerhebung

2.1 Methodik und Merkmale

Die Arbeitskostenerhebung im Produzierenden Gewerbe und im Dienstleistungsbereich wird seit 1957 durchgeführt, seit 1959 auf Grundlage einer EG-Verordnung. Die Periodizität war zunächst unregelmäßig alle 3 bis 4 Jahre. Seit den 80er Jahren wird die Arbeitskostenerhebung vierjährlich durchgeführt.

Zum Berichtskreis gehörte zunächst nur das Produzierende Gewerbe. Ab 1970 wurden sukzessive Teile des Dienstleistungsbereiches mit einbezogen. Aktuell gehören dazu die Abschnitte G (Handel; Instandhaltung und Reparatur von Kraftfahrzeugen und Gebrauchsgütern), H (Gastgewerbe), I (Verkehr und Nachrichtenübermittlung), J (Kredit- und Versicherungsgewerbe) und K (Grundstücks- und Wohnungswesen, Vermietung beweglicher Sachen, Erbringung von Dienstleistungen überwiegend für Unternehmen). Wegen fehlender aktueller Auswahlgrundlagen wurde der Abschnitt K aber 1996 und 2000 nicht erfasst. Für Unternehmen des Produzierenden Gewerbes ist die Befragungseinheit der Betrieb; Unternehmen des Dienstleistungsbereiches geben für jedes Bundesland, in dem sie vertreten sind, einen Erhebungsbogen ab, der zusammengefasste Angaben für alle Unternehmensteile in diesem Bundesland enthält.

Die Arbeitskostenerhebung wird als einfach geschichtete Stichprobe durchgeführt. Dabei erfolgt die Schichtung nach den 16 Bundesländern, 38 Wirtschaftsgruppen und 7 Beschäftigtengrößen-

3) Systematik der Wirtschaftszweige, Ausgabe 1979, Fassung für die Statistik im Produzierenden Gewerbe (SYPRO), Arbeitsunterlage, Statistisches Bundesamt 1989.

klassen. Für das Jahr 2000 liegen Angaben von gut 29 000 Unternehmen für ca. 40 000 Betriebe bzw. Unternehmensteile vor. Wie bei der GLS gilt auch hier das Prinzip des vergleichbaren Standardfehlers, das für kleinere Bundesländer größere Auswahlsätze erfordert. Analysen sind regional maximal auf der Ebene der Bundesländer möglich, da der Datensatz keine tiefere Regionalinformation enthält.

Das Ziel der Erhebung besteht darin, Daten zur Höhe und der Struktur der Kosten zu erhalten, die den Unternehmen durch die Beschäftigung von Arbeitnehmern entstehen. Dazu sind diese Kosten in etwa 40 Positionen unterteilt; im Produzierenden Gewerbe nach Arbeitern und Angestellten getrennt, im Dienstleistungsbereich nur für alle Arbeitnehmer insgesamt. Die Differenzierung nach Arbeitern und Angestellten im Produzierenden Gewerbe entfällt ab 2004. Die Angaben beziehen sich jeweils auf das ganze Erhebungsjahr bzw. auf das Geschäftsjahr, falls es vom Kalenderjahr abweicht. Knapp 90 % der Daten der Erhebung 2000 gelten tatsächlich für das Kalenderjahr 2000.

Zu den Kostenpositionen, die ausgewiesen werden, gehören u. a.:

- Sonderzahlungen
- Leistungen zur Vermögensbildung
- Zusätzliches Urlaubsgeld
- Zusätzliche Gehaltsfortzahlung im Krankheitsfall
- Pflichtbeiträge zur Sozialversicherung
- Beiträge zur Berufsgenossenschaft
- Betriebliche Rentenzahlungen
- Pensionsrückstellungen
- Beiträge für Direktversicherungen
- Kurzarbeitergeld
- Abfindungen bei Beendigung des Arbeitsverhältnisses
- Sonstige Sozialleistungen
- Wohnungsfürsorge
- Kantinenkosten, Essensgutscheine, Essensmarken
- Zuschüsse zum Mutterschaftsgeld
- Kosten für berufliche Aus- und Weiterbildung

Ferner sind für den Betrieb bzw. Unternehmensteil die Anzahl der Beschäftigten differenziert nach Vollzeit, Teilzeit und Auszubildenden, die Anzahl der bezahlten Arbeitsstunden, die Anzahl der effektiv geleisteten Arbeitsstunden, die Wochenarbeitszeit des häufigsten Vollzeit- und Teilzeitmodells und die Anzahl der durch Kurzarbeit oder Streik ausgefallenen Stunden angegeben. Schließlich ist ein Faktor zur Hochrechnung der Ergebnisse enthalten. Dieser ergibt sich als Quotient aus der Anzahl der Fälle in der Auswahlgrundlage in der Schicht und der Anzahl der Fälle in der Stichprobe in der Schicht.

Weitere Informationen zu Methodik und Merkmalen der Arbeitskostenerhebung findet man in Kaukewitsch und Heppt (1999) und in den Metadaten im Internetangebot der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder.

2.2 Datenzugang für Wissenschaftler

Die Einzeldaten der Erhebung 2000 stehen seit September 2005 zur On-Site-Nutzung an den Gastwissenschaftlerarbeitsplätzen an allen Standorten des FDZ der Statistischen Landesämter und mittels kontrollierter Datenfernverarbeitung zur Verfügung. Die Aufbereitung der Daten von 1996 und 1992 ist für Frühjahr 2006 geplant. Die Erhebung für 2004 wurde im Frühjahr 2005 von den Landesämtern durchgeführt. Die Prüfung und Plausibilisierung der Daten wird voraussichtlich bis Mitte 2006 abgeschlossen, so dass die Aufbereitung für das FDZ im Sommer 2006 erfolgen kann.

Literatur

Bechtel, Stephan, Mödinger, Patricia und Strotmann, Harald: Tarif- und Lohnstrukturen in Baden-Württemberg: Entwicklung und Einfluss der Tarifbindung auf Verdiensthöhe und -streuung, Statistische Analysen 7/2004, Statistisches Landesamt Baden-Württemberg.

Blinder, Alan S.: Wage Discrimination: Reduced Form And Structural Estimates, The Journal Of Human Resources VIII (1973), 436 – 455.

Fitzenberger, Bernd und Reize, Frank: Verteilung, Differentiale und Wachstum – Eine Verdienstanalyse für Westdeutschland auf Basis der Gehalts- und Lohnstrukturerhebung, Discussion Paper No. 02 – 71 (2002), ZEW Mannheim, <ftp://ftp.zew.de/pub/zew-docs/dp/dp0271.pdf>.

Frank-Bosch, Birgit: Verdienststrukturen in Deutschland: Methode und Ergebnisse der Gehalts- und Lohnstrukturerhebung 2001, Wirtschaft und Statistik 12 (2003), 1137 – 1151.

Groshen, Erica L.: The Structure Of The Female/Male Wage Differential, The Journal Of Human Resources 26 (1991), 455 – 472.

Jirjahn, Uwe und Stephan, Gesine: Betriebliche Sonderzahlungen – Theoretische Überlegungen und empirische Befunde. In: B. Frick, R. Neubäumer und W. Sesselmeier (Hrsg.), Die Anreizwirkungen betrieblicher Zusatzleistungen, Hampp u. a. München S. 33 – 67 Reihe/Serie: Organisationsökonomie humaner Dienstleistungen Nr. 06, 1999.

Kaukewitsch, Peter und Heppt, Ehrenfried: Arbeitskosten im Produzierenden Gewerbe und ausgewählten Dienstleistungsbereichen, Wirtschaft und Statistik 2 (1999), 116 – 128.

von Kulmiz, Leontine: Lohndiskriminierung von Frauen. Eine Analyse mit der Gehalts- und Lohnstrukturerhebung 1990, Wirtschaft und Statistik 5 (2001), 406 – 415.

Oaxaca, Ronald: Male-Female Wage Differentials In Urban Labor Markets, International Economic Review 14 (1973), 693 – 709.

Stephan, Gesine: Eine empirische Analyse der Lohnunterschiede zwischen Frauen und Männern, Statistische Monatshefte Niedersachsen 1 (1997), 5 – 11.

Stephan, Gesine und Gerlach, Knut: Firmenlohndifferenziale und Tarifverträge – eine Mehrebenenanalyse, Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 36, H. 4 (2003), 525 – 538.

Autoren- und Kontaktangaben:

Dr. Hans-Peter Hafner, FDZ der Statistischen Landesämter, Statistisches Landesamt Hessen,
E-Mail: forschungsdatenzentrum@statistik-hessen.de

Erste Scientific-Use-Files aus den Wirtschaftsstatistiken

Roland Sturm, Rainer Lenz

Das Projekt „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ (FAWE) fand Wege, wie für Firmendaten eine faktische Anonymisierung gelingen kann. Gefördert vom Bundesministerium für Bildung und Forschung haben die statistischen Ämter damit gemeinsam mit der Wissenschaft für Unternehmens- und Betriebsdaten das so genannte Wissenschaftsprivileg des Bundesstatistikgesetzes mit Leben gefüllt.

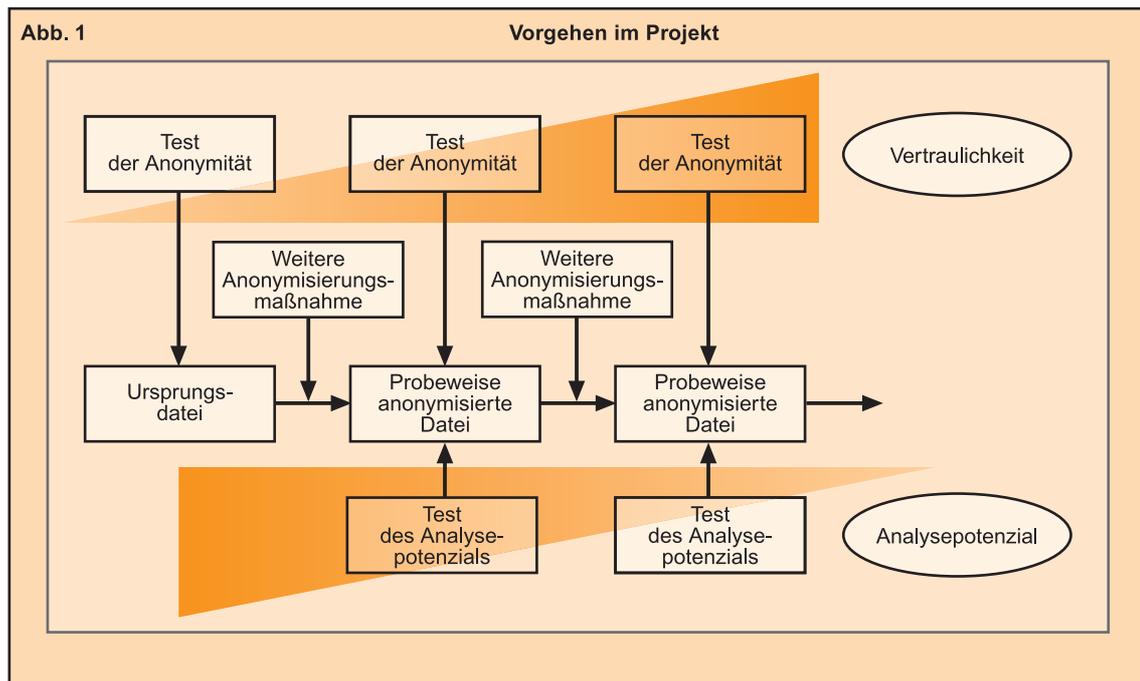
Faktisch anonymisierte Einzeldaten müssen zwei gleichrangigen Ansprüchen gerecht werden: Einerseits verlangt das Bundesstatistikgesetz, dass ein ausreichender Schutz der Einzelangaben gewährleistet ist; andererseits müssen die Analysemöglichkeiten der anonymisierten Daten so weit wie möglich erhalten bleiben. Beide Ziele sind sowohl für die Nutzer als auch für die statistischen Ämter von großer Bedeutung. In diesem Beitrag werden die wesentlichen Ergebnisse des Projektes vorgestellt: Zum einen Empfehlungen für die Anonymisierung wirtschaftsstatistischer Einzeldaten und zum anderen erste Datenangebote, die bereits zur Laufzeit des Projektes entstanden sind.

Der Anspruch, den die Initiatoren des Projektes an ihre Arbeit hatten, ging über die Schaffung eines konkreten Datenangebotes für bestimmte Erhebungen der statistischen Ämter hinaus. Das Kompendium, das mit dem kürzlich veröffentlichten „Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten“ vorgelegt wird, liefert den Ämtern und allen weiteren interessierten Datenanbietern Hilfen für eine sinnvolle und effiziente Anonymisierung. Es bietet die Grundlage für die Methodenauswahl, die Anwendung der Methoden und die Prüfung ihrer Wirkungsweise hinsichtlich künftiger Anonymisierungen.

Das Projekt FAWE – Ziel und Vorgehen

Das Projekt „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ beschäftigte sich mit der Frage, wie eine faktische Anonymisierung von Unternehmens- und Betriebsdaten gelingen kann. Gemeinsam mit der Wissenschaft haben die statistischen Ämter das so genannte Wissenschaftsprivileg des Bundesstatistikgesetzes (BStatG) umgesetzt, nach dem für die Durchführung wissenschaftlicher Vorhaben faktisch anonymisierte Mikrodaten an Hochschulen und sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung übermittelt werden dürfen.

Faktisch anonymisierte Einzeldaten müssen zwei Eigenschaften erfüllen: Das BStatG verlangt einen ausreichenden Schutz der Informationen, die die Auskunftgebenden der Statistik überlassen. Gleichzeitig nutzen die Informationen für die wissenschaftliche Analyse nur, wenn damit valide Ergebnisse produziert werden können.



Im Projekt FAWE wurden die beiden Aspekte „Schutzwirkung“ und „Analysepotenzial“ umfangreich untersucht. Abbildung 1 zeigt schematisch das Vorgehen bei der Anonymisierung einer Datei. Zunächst werden die Ursprungsdaten auf ihre Schutzbedürftigkeit hin getestet. Die Entfernung direkter Identifikatoren wie Name und Adresse eines Unternehmens, Dateninkompatibilitäten zu externem Zusatzwissen, die Beschränktheit des Zusatzwissens auf wenige sogenannte Überschneidungsmerkmale und Eigenschaften der jeweiligen Statistik (z. B. Stichprobe, älteres Erhebungsjahr der Daten) entfalten bereits ohne eine Behandlung der Originalwerte einen Schutz. Reicht diese „natürliche“ Schutzwirkung nicht aus, um die Daten aller betroffenen Einheiten einer Erhebung ausreichend zu schützen, so werden im nächsten Schritt Anonymisierungsmaßnahmen auf die Daten angewendet. Neben den bereits bisher in den statistischen Ämtern eingesetzten traditionellen informationsreduzierenden Anonymisierungsmaßnahmen werden im Projekt auch Auswirkungen datenverändernder Verfahren untersucht.¹⁾ Im schematischen Vorgehen wird zunächst eine schwache Dosierung der Anonymisierung gewählt, um möglichst viel des Analysepotenzials in den Daten zu erhalten. Die so gewonnene probeanonymisierte Datei wird auf ihre Schutzwirkung und auf ihr Analysepotenzial untersucht. Abhängig vom Ergebnis kann ein weiterer Anonymisierungsschritt (z. B. stärkere Dosierung) erfolgen und wiederum auf seine Auswirkung geprüft werden. Bei der geschilderten Steigerung der Dosierung der Anonymisierungsmaßnahmen ist zu erwarten, dass die Schutzwirkung steigen und das Analysepotenzial abnehmen wird. Es muss entschieden werden, wann Anonymisierungsmaßnahmen einen ausreichenden Schutz erzeugt haben, der die gewünschte Anonymität gewährleistet.

In diesem Beitrag werden die wesentlichen Ergebnisse des Projektes FAWE in Kurzform zusammengefasst. Dies sind zum einen Empfehlungen für die Anonymisierung wirtschaftsstatistischer Einzeldaten (Abschnitt 2) und zum anderen erste Datenangebote, die bereits zur Laufzeit des Projektes entstanden sind (Abschnitt 3).

1) Eine Übersicht über Anonymisierungsmethoden und die Auswahl der im Projekt untersuchten Methoden siehe Höhne, J.: Methoden zur Anonymisierung wirtschaftsstatistischer Einzeldaten, in: Gnos, Ronning (Hrsg.): Anonymisierung wirtschaftsstatistischer Einzeldaten, Forum der Bundesstatistik, 2003, S. 69 – 94.

2. Projektergebnisse I: Empfehlungen zur Anonymisierung

2.1 Schutz der Vertraulichkeit

Das Wissenschaftsprivileg für den Zugang zu faktisch anonymisiertem Einzelmaterial der statistischen Ämter ist im Bundesstatistikgesetz (BStatG) formuliert. Gemäß dem Gebot des § 16 Abs. 6 BStatG können Einzeldaten dann an wissenschaftliche Einrichtungen übermittelt werden, wenn sie nur mit einem „unverhältnismäßig hohen Aufwand“ zugeordnet (reidentifiziert) werden können. Dieses Unverhältnismäßigkeitsgebot ist gemeint, wenn von faktischer anstelle absoluter Anonymität gesprochen wird.

Im Rahmen des Projektes FAWE wurde ein „Konzept zur Beurteilung der Schutzwirkung von faktischer Anonymisierung“ entwickelt. Um das Wissenschaftsprivileg des § 16 (6) BStatG beim Datenzugang nicht unnötig einzuschränken, darf eine faktische Anonymisierung nur auf realistische Gefährdungsszenarien abstellen und nicht dazu führen, dass unnötig hohe Schutzmaßnahmen aufgebaut werden. Zur Operationalisierung, was bei der Reidentifikation von Einzeldaten ein „unverhältnismäßig hoher Aufwand“ ist, werden im genannten Konzept realitätskonforme Annahmen für einen Datenangriff zugrunde gelegt. Aus dem Rationalkalkül eines potenziellen Datenangreifers²⁾ kann abgeleitet werden: Eine Verletzung der faktischen Anonymität wird nur dann erfolgen, wenn eine mögliche Reidentifikation für einen Datenangreifer als nutzbringend einzustufen ist. Die Herstellung einer nutzbringenden Reidentifikation wäre für einen Datenangreifer erwägenswert, wenn er abschätzen könnte, dass der (erwartete) Ertrag eines Reidentifikationsversuchs dessen (erwarteten) Aufwand übersteigt. Kosten und Nutzen sind davon abhängig, wie „sicher“ ein Datenangreifer brauchbare Informationen enthüllen kann. In der Praxis wird sich ein Datenangreifer mit zahlreichen Problemen konfrontiert sehen:

- mit der Existenz von Dateninkompatibilitäten zwischen Datensätzen aus dem Zusatzwissen und amtlichen Datensätzen,
- mit der Unkenntnis, ob eine gesuchte Einheit in der amtlichen Erhebung enthalten ist (insbesondere im Falle von Stichprobenerhebungen),
- mit Unsicherheiten über die Richtigkeit von Zuordnungsversuchen und
- mit Unsicherheiten über die Qualität von enthüllten Informationen.

Zu Beginn der Untersuchung der Schutzwirkung probeweise anonymisierter Daten sollte eine ausführliche Recherche über das mögliche Zusatzwissen eines potenziellen Datenangreifers durchgeführt werden. Bereits hier können kritische, bei Datenangriffen besonders gefährdete Bereiche in den Daten aufgedeckt werden. Diese sind zum Teil den Fachleuten schon vor der Recherche bekannt (z. B. niedrige Besetzungszahlen in Tabellen einer Fachserie). Desweiteren sollten sich die Fachleute der Erhebung und der Anonymisierung zusammensetzen und sich auf Risikoschwellen verständigen.

In einem nächsten Schritt sollte versucht werden, eine Datenbank mit dem relevanten Zusatzwissen für Massenfischzüge aufzubauen. Da im Allgemeinen keine gemeinsamen Identifikato-

2) Zum Konzept des Datenangreifers und des Datenangriffs vgl. Höhne, J., Sturm, R., Vorgrimler, D.: Konzept zur Beurteilung der Schutzwirkung faktischer Anonymisierung in: Wirtschaft und Statistik Heft 4 2003, S 287 – 292

ren zwischen Daten verschiedener Quellen vorliegen, ist hier auch seitens der Datenanbieter mit viel Aufwand zu rechnen.³⁾ In den meisten Fällen liegen in beiden Quellen zumindest Namen und/oder Adressen vor, die allerdings stark voneinander abweichen können.

Ist Zusatzwissen in Form einer Datenbank vorhanden, so kann ein Massenfischzug simuliert werden. Ein Programm hierzu kann über das Forschungsdatenzentrum des Statistischen Bundesamtes bezogen werden.⁴⁾ Dieses Programm wäre selbstverständlich für einen potenziellen Datenangreifer unbrauchbar, da dieser keine Möglichkeit hat, die durch das Programm getroffenen Zuordnungen zu überprüfen.

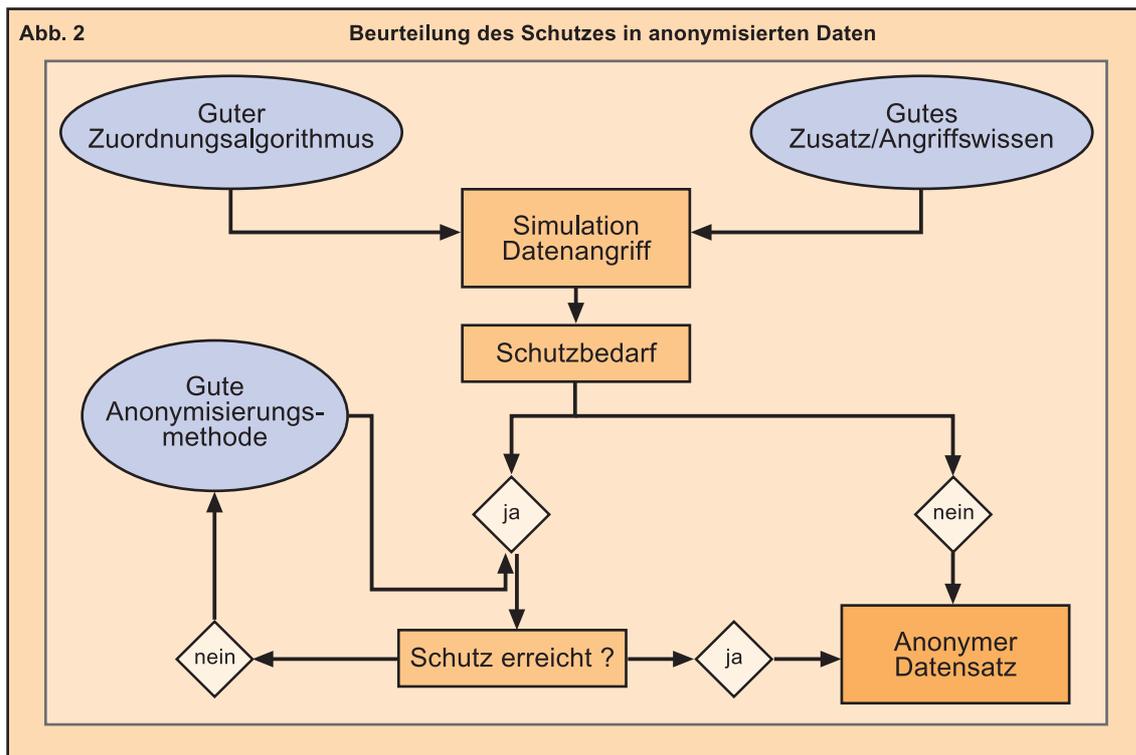
Das Vorgehen bei der Anonymisierung einer bestimmten Erhebung – mit besonderem Fokus auf die gefährdeten Bereiche – besteht oftmals aus einer Mischung aus informationsreduzierenden und datenverändernden Methoden. Informationsreduzierende Methoden wie die Entfernung von Merkmalen, die für den Datennutzer nicht von Belang sind oder eine (weitere) Vergrößerung kategorialer Überschneidungsmerkmale in der Form, dass die den Wissenschaftler interessierenden Teilmassenauswertungen weiterhin möglich sind, sollten bevorzugt angewendet werden, da hier Informationen nicht verfremdet, sondern lediglich unterdrückt werden. Auf diese Weise können mögliche, für den Datenangreifer nahezu unverzichtbare Blockmerkmale entschärft werden, da diese nach der Vergrößerung die Daten feiner partitionieren. In Datenbereichen mit verhältnismäßig dichter Besetzung kann eine Anonymisierung allein mit informationsreduzierenden Methoden (wie z. B. bei Unternehmen der Einzelhandelsstatistik mit weniger als 50 Beschäftigten, siehe Abschnitt 3.2) oder mit geringfügiger Datenveränderung (wie z. B. bei Unternehmen der Kostenstrukturerhebung im Verarbeitenden Gewerbe mit weniger als 50 Beschäftigten, siehe Abschnitt 3.1) gelingen. In Bereichen mit dünner Besetzung sind dagegen datenverändernde Verfahren in der Regel unvermeidlich, wie z. B. bei marktführenden Unternehmen der Umsatzsteuerstatistik in bestimmten Branchen (siehe Abschnitt 3.3).

Hat man sich bei den datenverändernden Anonymisierungsverfahren auf eine bestimmte Verfahrensgruppe verständigt, so müssen im Folgenden die Parameter des Verfahrens ausbalanciert und Datenangriffe so oft simuliert werden, bis der gewünschte Grad an Anonymisierung erreicht und eine a priori vorgegebene obere Risikoschwelle unterschritten wird. Abbildung 2 beinhaltet zusammengefasst das Vorgehen bei der Beurteilung des Schutzes in anonymisierten Daten. Wie im ersten Abschnitt bereits erwähnt wurde, sollten diese Untersuchungen nicht isoliert von der Bewertung des Analysepotenzials durchgeführt werden. In die Parameter der Anonymisierung sollten simultan Untersuchungen hinsichtlich des Analysepotenzials eingehen.

2.2 Analysepotenzial

Niemandem ist damit gedient, anonymisierte Daten anzubieten, die die Nutzer nicht einsetzen können oder wollen. Daher wurde das Projekt FAWE von Datenanbietern und Datennutzern gemeinsam durchgeführt. Der Projektpartner IAW (Institut für Angewandte Wirtschaftsforschung)

3) Siehe Lenz, R., Doherr, T., Vorgrimler, D.: „Simulation of a database cross match“ in: Proceedings of the European Conference on Quality and Methodology in Official Statistics (Q2004), Mainz. – 4) Eine theoretische Beschreibung samt Anwendungsbeispiel des Programmes findet sich in Lenz, R.: „Measuring the disclosure protection of microaggregated business microdata – An analysis taking the example of German Structure of Costs Survey“, erscheint in: Journal of Official Statistics.



hatte die Aufgabe, die Nutzeranforderungen im Projekt zu vertreten. Um die Bedürfnisse der Nutzer besser kennenzulernen, wurden eine Nutzerbefragung und zwei Nutzerworkshops durchgeführt. Die Ergebnisse geben einen guten Einblick in die Bedürfnisse der zukünftigen Anwender anonymisierter Daten. Diese Erkenntnisse wurden im Verlauf der Projektarbeiten aufgegriffen und berücksichtigt. Die starke Nutzerbeteiligung wurde durch die Einrichtung eines Wissenschaftlichen Begleitkreises aus empirischen Wirtschafts- und Sozialforschern, der die gesamten Projektarbeiten begleitet hat, beim IAW verankert. Als Novum in der Anonymisierungsforschung wird der Begriff des Analysepotenzials bzw. seine Veränderung infolge der Anonymisierung systematisiert und umfassend operationalisiert. Auf Basis dieser Arbeiten wurde im Projekt das Analysepotenzial der anonymisierten Daten beurteilt und ein Raster an Empfehlungen für die Bewertung künftiger Anonymisierungen bereitgestellt. Darüberhinaus werden Anregungen an die Datenanbieter gegeben, wie sie auf Unsicherheiten und Vorbehalte der Datennutzer eingehen und diese beim sachgerechten Einsatz anonymisierter Daten unterstützen können. Ausführliche Analysen hierzu finden sich im Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten.⁵⁾ An dieser Stelle sei auch auf den Beitrag von Martin Rosemann in diesem Band verwiesen.

Die Einschränkungen des Analysepotenzials sowohl durch informationsreduzierende als auch durch datenverändernde Anonymisierungsverfahren sind nicht objektiv bewertbar. Vielmehr müssen die Auswirkungen solcher Verfahren auf das Analysepotenzial und die Rückschlüsse für ihren Einsatz in einem diskursiven Prozess zwischen Datenhaltern und Datennutzern geklärt werden. Sie lassen sich wie folgt systematisieren:⁶⁾

5) Ronning, G. et al.: „Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten“, in: Statistik und Wissenschaft, Band 4, Statistisches Bundesamt. – 6) Ronning, G., Rosemann, M.: „Ansätze zur Operationalisierung des Analysepotenzials bei anonymisierten Daten“, Arbeitspapier des Projektes FAWE, 2003.

- Maßzahlorientierter Ansatz
- Anwendungsorientierter Ansatz
- Theorieorientierter Ansatz

Während die beiden ersten Ansätze auf einer empirischen Herangehensweise beruhen, orientiert sich der dritte Ansatz an der theoretischen Ableitung der Auswirkung von Anonymisierungsverfahren auf Analyseergebnisse.

Die Idee des *maßzahlorientierten* Ansatzes besteht in der Abschätzung der Veränderung des Analysepotenzials durch die Abweichung geeigneter Maße, mit denen üblicherweise empirische Verteilungen charakterisiert werden. In der Literatur werden eine Reihe von Abweichungsmaßen vorgestellt.⁷⁾ Dabei können für metrische Merkmale folgende Arten von Maßen unterschieden werden:

- Abweichungen der einzelnen Werte
- Abweichungen der Mittelwerte und Varianzen (univariate Verteilungen)
- Abweichungen von Konzentrationsmaßen (univariate Verteilungen)
- Abweichungen von Zusammenhangsmaßen (multivariate Verteilungen)
- Abweichungen höherer Momente

Der *anwendungsorientierte* Ansatz verfolgt die Idee, die Auswirkungen verschiedener Anonymisierungsverfahren auf das Analysepotenzial von Einzeldaten danach zu bewerten, ob und inwiefern sich die Ergebnisse verschiedener Analysen verändern. Dies kann in der Praxis dadurch umgesetzt werden, dass ein repräsentativer Querschnitt an verschiedenen Fragestellungen mit unterschiedlichen statistischen Methoden untersucht wird. Dabei können sowohl die inhaltlichen Fragestellungen als auch die verwendeten Methoden stets nur beispielhaft sein. Der anwendungsorientierte Ansatz stellt damit gewissermaßen eine Weiterentwicklung des maßzahlorientierten Ansatzes dar, indem er die Zahl der relevanten Entscheidungskriterien auf die Ergebnisse statistischer und ökonomischer Analysen ausdehnt. Die Entscheidungskriterien können jedoch nie vollständig sein, da nicht überblickt werden kann, welche möglichen Fragestellungen in der Zukunft mit einem Datenbestand untersucht werden. Einbezogen werden sollten möglichst unterschiedliche Analysemethoden, insbesondere deskriptive Auswertungen und ökonomische Schätzungen. Die inhaltlichen Fragestellungen hängen dabei von den Informationen im Datensatz, also von den im Datensatz vorhandenen Merkmalen ab. Im Unterschied zum maßzahlorientierten Ansatz können mit Hilfe dieses Ansatzes sowohl die inhaltlichen Einschränkungen an Analysemöglichkeiten durch die Informationsreduktion im Datensatz als auch die durch Datenveränderung entstehenden Veränderungen der Analyseergebnisse beurteilt werden.

Während die beiden empirischen Ansätze davon ausgehen, dass eine größtmögliche Erhaltung des Analysepotenzials auch gleichzeitig eine größtmögliche Gleichheit der empirischen Ergebnisse von Analysen bzw. die weitgehende Erhaltung von Verteilungsmaßen bedeutet, besteht der theorieorientierte Ansatz darin festzustellen, ob es Auswirkungen der Anonymisierungsverfahren auf Eigenschaften von Schätzern (z. B. in ökonomischen Modellen) oder deskriptive

7) Siehe z. B. Domingo-Ferrer, J., Mateo, J., Torres, A.: „Concepts for the evaluation of anonymized data“, in: Gnos, R., Ronning, G. (Hrsg.), Anonymisierung wirtschaftsstatistischer Einzeldaten, Forum der Bundesstatistik, Band 42, 100 – 110, 2003. Siehe auch Fußnote 6.

statistische Maße gibt bzw. ob die Schätzer erwartungstreu oder korrigierbar sind. Hierzu wurden die Auswirkungen datenverändernder Verfahren für verschiedene Modelle oder Maße theoretisch abgeleitet.⁸⁾ Der Vorteil des theorieorientierten Ansatzes besteht darin, dass eine objektive Bewertung der datenverändernden Verfahren dahingehend erfolgen kann, ob die Verfahren theoretisch zu keiner Veränderung bestimmter Analyseergebnisse führen bzw. erwartungstreu sind oder sich die Erwartungstreue durch Korrekturverfahren wiederherstellen lässt. Allerdings ist eine theoretische Untersuchung einer Vielzahl von Modellen und deskriptiver Maße erforderlich, um eine umfassende Bewertung auf Basis dieses Ansatzes abgeben zu können. Zudem können selbst korrigierte Verfahren, die in der Theorie zu erwartungstreuen Schätzern führen, in der Praxis bei endlichen Stichproben einen verzerrten Schätzer erzeugen.

3 Projektergebnisse II: Datenangebote

3.1 Kostenstrukturerhebung im Verarbeitenden Gewerbe

Inhaltlich liefern die Kostenstrukturerhebungen im Verarbeitenden Gewerbe (KSE) die umfassendsten Informationen zu den Unternehmen im Bereich der Statistik im Produzierenden Gewerbe. Sie dienen als Ausgangspunkt für vielfältige Strukturuntersuchungen, nicht nur in Politik und Verwaltung, sondern auch in der Wirtschaft und ihren Verbänden sowie in der Wissenschaft und vielen anderen gesellschaftlichen Gruppierungen. Die Informationen der KSE bilden darüber hinaus eine unentbehrliche Datengrundlage für die Volkswirtschaftlichen Gesamtrechnungen. Hier werden die Ergebnisse vor allem für die Berechnung der Wertschöpfung und ihrer Komponenten nach Wirtschaftsbereichen im Rahmen der Entstehungsrechnung herangezogen; schließlich liefern sie auch wichtige Informationen für die Input-Output-Rechnungen. Im Rahmen der Statistiken im Produzierenden Gewerbe bilden die Kostenstrukturstatistiken u. a. eine Grundlage für die Gewichtung von Produktionsindizes. Die Kostenstrukturerhebung enthält zahlreiche Informationen zu Beschäftigtengrößen, Umsätzen und Kostenstruktur der Unternehmen.

Die Kostenstrukturerhebung im Verarbeitenden Gewerbe des Jahres 1999 – reduziert auf Unternehmen mit 20 bis einschließlich 249 Beschäftigten – erfasst als hochrechnungsfähige Stichprobe von ca. 38 % etwa 13 000 Unternehmen. Die Befragung erfolgt zentral durch das Statistische Bundesamt im Wege der Selbstaussfüllung durch die Unternehmen. Die in der Stichprobe gewonnenen Ergebnisse werden auf die Gesamtheit der Unternehmen zwischen 20 und 249 Beschäftigten hochgerechnet.⁹⁾

Bei der Anonymisierungsstrategie wurde – den Anregungen der Nutzer folgend – ein stärkeres Gewicht auf die Behandlung der qualitativen Merkmale gelegt.¹⁰⁾ In einem ersten Schritt wurde

8) Siehe Ronning, G.: „Randomized Response and the Binary Probit Model“, in: Economics letters, Bd. 86, 221 – 228, 2005. Siehe auch Lechner, S., Pohlmeier, W.: „Schätzung ökonometrischer Modelle auf der Grundlage anonymisierter Daten“, in: Gnos, R., Ronning, G. (Hrsg.): Anonymisierung wirtschaftsstatistischer Einzeldaten, Forum der Bundesstatistik, Bd. 42, 2003. – 9) Für eine detaillierte Beschreibung siehe „Beschreibung der Kostenstrukturerhebung im Verarbeitenden Gewerbe“, Zwischenbericht zum Projekt „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“, Juni 2002, Wiesbaden. – 10) Hier ergibt sich eine Parallele zur faktischen Anonymisierung der Einkommensteuerstatistik, auch in diesem Fall sprach sich der Wissenschaftliche Beraterkreis für einen Erhalt der quantitativen zu Lasten der qualitativen Merkmale aus, siehe Merz, J., Vorgrimler, D., Zwick, M.: „Faktisch anonymisiertes Mikrodatenfile der Lohn- und Einkommensteuerstatistik 1998“, in: Wirtschaft und Statistik 10/2004, S. 1079 ff.

auf das ursprünglich im Merkmalskanon vorhandene Merkmal „Tätige Inhaber“ verzichtet, da es sich im Laufe der Projektarbeiten als sehr reidentifikationsgefährdend und für wissenschaftliche Analysen als wenig wertvoll herausgestellt hat.

Besonders geeignet für Reidentifikationen sind regionale Angaben. Der Erhalt solcher Merkmale in einem Scientific-Use-File stellt daher für die Anonymisierung ein schwieriges Unterfangen dar. Bereits zu Beginn des Projektes wurde die Möglichkeit ausgeschlossen, einen Scientific-Use-File zu erstellen, der administrative Gebietsangaben auf der Ebene der Bundesländer oder gar einer tieferen Gliederungsebene enthält. Da aber die Auswertung nach Regionen einen wichtigen Analysebereich darstellt, wurde nach alternativen Möglichkeiten gesucht und dies vor dem Hintergrund, gleichzeitig auf datenverändernde Maßnahmen bei den quantitativen Merkmalen weitestgehend verzichten zu wollen. Als erste Möglichkeit wurde der administrative Gebietsschlüssel durch den nichtadministrativen siedlungsstrukturellen Kreistyp BBR9 und den siedlungsstrukturellen Regionstyp BBR3 ersetzt. Diese Vergrößerung der Regionalinformation hatte eine sehr deutliche Schutzwirkung.¹¹⁾ Allerdings sprach sich der Wissenschaftliche Begeleitkreis dafür aus, anstelle dieser nicht-administrativen Schlüssel eine Ost-West-Klassifizierung einzuführen. Diese zweite Möglichkeit wurde schließlich in den Scientific-Use-File aufgenommen.

Die Daten der Kostenstrukturerhebung im Verarbeitenden Gewerbe werden nach der Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ93), auf der Vierstellerebene (Klasse) erhoben und aufbereitet. Diese Klassifikation ist von der europäischen Klassifikation NACE Rev.1 abgeleitet, die aufgrund der NACE-Verordnung des Rates der Europäischen Gemeinschaften seit 1995 in allen Mitgliedstaaten der Europäischen Union sowohl für die Erhebung als auch für die Darstellung der statistischen Daten anzuwenden ist.¹²⁾ Das Kodierungssystem der WZ93 unterscheidet zwischen Abschnitten (Buchstaben A-Q), Unterabschnitten (Buchstaben AA-QA), Abteilungen (Zweisteller), Gruppen (Dreisteller), Klassen (Viersteller) und Unterklassen (Fünfsteller). Der Wirtschaftsbereich „Verarbeitendes Gewerbe sowie Bergbau und Gewinnung von Steinen und Erden“ erstreckt sich über die Abschnitte C und D bzw. – in der numerischen Gliederung – über die Abteilungen 10 bis 37. Im Projekt haben sich Datenanbieter und Datennutzer darauf verständigt, bei dem hierarchischen Merkmal WZ93 die Gliederungstiefe 2 (Zweistellerebene) nicht zu unterschreiten, da hierdurch zum einen eine beachtliche Schutzwirkung (beispielhafte Berechnungen finden sich in (Lenz/Sturm/Vorgrimler)¹³⁾ und zum anderen nach Einschätzung der beteiligten Wissenschaftler für einen Scientific-Use-File eine ausreichende Breite an Analysemöglichkeiten¹⁴⁾ erhalten wird.

In den Veröffentlichungen der statistischen Ämter werden aufgrund von Geheimhaltungsaspekten die Ergebnisse einiger Wirtschaftsabteilungen nicht veröffentlicht. Es handelt sich dabei um Unternehmen der Abteilungen 10, 11, 14, 16, 23, 30, 32, 35 und 37 der WZ93. Bei den im Projekt durchgeführten Simulationen hat sich bestätigt, dass diese Abteilungen neben den Abteilungen 15, 17, 18, 19, 22 und 34 größerer Geheimhaltung bedürfen. Um diese kritischen Abteilungen im Scientific-Use-File belassen und weitgehend auf datenverändernde Verfahren bei den

11) Siehe Lenz, R., Vorgrimler, D.: „Matching German Turnover Tax Statistics“, Diskussionspapier Nr. 4, Forschungsdatenzentren des Bundes und der Länder. – 12) Für neuere Erhebungen ab dem Jahr 2003 gilt mit dem Branchenschlüssel WZ 2003 wiederum eine neue Klassifikation. – 13) Lenz, R., Sturm, R., Vorgrimler, D.: „Maße für die faktische Anonymität von Mikrodaten“ in *Wirtschaft und Statistik* 6/2004, S. 621 – 638. – 14) Siehe Rosemann, M., Vorgrimler, D., Lenz, R.: „Erste Ergebnisse faktischer Anonymisierung wirtschaftsstatistischer Einzeldaten“ in *Allgemeines Statistisches Archiv* Band 88 (1), 2004, S. 73 – 99.

quantitativen Merkmalen verzichten zu können, werden die Abteilungen 10 (Kohlenbergbau, Torfgewinnung), 11 (Gewinnung von Erdöl und Erdgas, Erbringung damit verbundener Dienstleistungen) und 14 (Gewinnung von Steinen und Erden, sonstiger Bergbau) zum Abschnitt C, die Abteilungen 15 (Ernährungsgewerbe) und 16 (Tabakverarbeitung) zum Unterabschnitt DA, die Abteilungen 17 (Textilgewerbe) und 18 (Bekleidungs-gewerbe) zum Unterabschnitt DB, die Abteilungen 21 (Papiergewerbe) und 22 (Verlags- und Druckgewerbe, Vervielfältigung) zum Unterabschnitt DE, die Abteilungen 30 (Herstellung von Büromaschinen, Dv-Geräten und -einrichtungen) und 31 (Herstellung von Geräten der Elektrizitätserzeugung, -verteilung u. ä.) zum Unterabschnitt DL sowie die Abteilungen 34 (Herstellung von Kraftwagen und Kraftwagenteilen) und 35 (Sonstiger Fahrzeugbau) zum Unterabschnitt DM zusammengefasst. Bei den Abteilungen 19 (Ledergewerbe) und 23 (Kokerei, Mineralölverarbeitung, Herstellung von Brutstoffen) wurde das Merkmal WZ93 unterdrückt. Außerdem wurde die Abteilung 37 (Recycling) aus inhaltlichen und aus Geheimhaltungsgründen herausgenommen. Obwohl die Abteilungen 32 (Rundfunk-, Fernseh- u. Nachrichtentechnik) und 33 (Medizin-, Mess-, Steuer- u. Regelungstechnik, Optik) ebenfalls zum Unterabschnitt DL zu zählen sind, werden Sie im Datensatz separat aufgeführt, da hier die Weitergabe der Zweisteller aus Sicht des Datenschutzes unbedenklich ist.

Die im Datensatz verbleibenden 30 quantitativen Merkmale wurden eindimensional für jedes Merkmal separat mikroaggregiert.¹⁵⁾ Eine Auflistung findet sich in Anhang 1. Bei dieser Variante der Mikroaggregation werden zunächst die Merkmalsausprägungen je Merkmal absteigend sortiert. Dann werden tripelweise (aus den Merkmalsausprägungen dreier benachbarter Merkmalsträger) die Durchschnittswerte ermittelt, die Originalwerte durch diese Durchschnittswerte ersetzt und wieder an die ursprüngliche Position zurücksortiert. Falls die Anzahl der Merkmalsträger nicht durch die Zahl Drei teilbar ist, so ist am Ende der absteigend sortierten Liste von Merkmalsausprägungen auch die Bildung einer Gruppe aus vier oder fünf Merkmalsträgern zulässig. Damit ist jede Merkmalsausprägung bei mindestens drei Merkmalsträgern vorhanden. Das hier skizzierte Verfahren ist für das Ziel einer möglichst vielseitigen Datennutzung, sowohl für deskriptive als auch für ökonometrische Auswertungen, das schonendste Verfahren innerhalb der Klasse der Mikroaggregationsverfahren.

Eine Liste aller im Datensatz enthaltenen Merkmale findet sich in Anhang 1.¹⁶⁾

3.2 Jahresherhebung im Einzelhandel

Die Ergebnisse der Jahresherhebung im Einzelhandel liefern wirtschaftspolitisch bedeutsame Informationen über die Struktur, Rentabilität und Produktivität der Unternehmen. Neben der Ermittlung des Rohertrages und der Bruttowertschöpfung können gute Schätzungen für die Vorratsveränderungen in der Wirtschaft gemacht werden. Aus konjunkturpolitischer Sicht sind sowohl die Beschäftigungssituation als auch die Lohn- und Gehaltsstrukturen interessant. Auf der betriebswirtschaftlichen Ebene lässt sich die Entwicklung der Arbeitsintensität und -produktivität überprüfen. Weiterhin gilt die Veränderungsrate der Investitionen als wichtiger Indikator für die längerfristige Umsatzerwartung eines Unternehmers. Eine Auflistung aller Merkmale des Scientific-Use-Files findet man im Anhang 2.

15) Zur Methode der Mikroaggregation siehe Fußnote 1. – 16) Eine ausführliche Beschreibung des Scientific-Use-Files findet sich in Lenz, R., Vorgrimler, D., Rosemann, M.: „Ein Scientific-Use-File der Kostenstrukturerhebung im Verarbeitenden Gewerbe“ in *Wirtschaft und Statistik* 2/2005, 91 – 96.

Die jährliche Einzelhandelsstatistik erfasst für das Jahr 1999 23 510 Unternehmen in der Stichprobe. Diese repräsentieren unter Einbezug der Hochrechnungsfaktoren knapp 300 000 Unternehmen des Einzelhandels mit einem Umsatz von rund 300 Mrd. Euro. Betrachtet man nur die Monatsmelder¹⁷⁾, so repräsentieren knapp 14 500 Unternehmen in der Stichprobe über 110 000 Unternehmen der Einzelhandelsstatistik, die einen Umsatz von rund 285 Mrd. Euro tätigen. Aufgrund der relativ hohen Anzahl von Unternehmen mit weniger als 50 Beschäftigten, kann für diese eine faktische Anonymisierung nahezu ohne datenverändernde Verfahren erreicht werden.

3.2.1. Scientific-Use-File für Unternehmen unter 50 Beschäftigte

Die Daten wurden formal anonymisiert, d. h. direkte Identifikatoren wie Name und Anschrift wurden entfernt. Umsatzdifferenzierungen nach Tätigkeiten bzw. Produkten wurden herausgenommen, da die Information über die Nebentätigkeiten eines Unternehmens die Reidentifikationsgefahr stark ansteigen lässt. Dies konnte anhand der kommerziell zugänglichen MARKUS-Datenbank festgestellt werden. Die Information der Haupttätigkeit eines Unternehmens bleibt durch die Kenntnis des Wirtschaftszweigs erhalten.

Das Merkmal „Anzahl der rechtlich unselbständigen örtlichen Einheiten des Unternehmens“ (Filialen) kann von einem Datenangreifer als Überschneidungsmerkmal benutzt werden. Daher wurden für die risikogefährdeten größeren Unternehmen bestimmte Kategorien für die Anzahl ihrer Filialen gebildet. Bei der Regionalangabe wurde auf Wunsch des Wissenschaftlichen Begleitkreises analog zum Vorgehen bei der Erstellung der Scientific-Use-Files der Kostenstrukturerhebung 1999 und der Umsatzsteuerstatistik 2000 die Ost-West-Klassifizierung verwendet.¹⁸⁾ Die Wirtschaftszweigklassifikation wurde auf den Dreisteller vergrößert. Ausnahme bildet der Dreisteller „524“ *Sonstiger Facheinzelhandel (in Verkaufsräumen)* der aufgrund seiner großen Besetzungszahl auf der Vierstellerebene angegeben werden kann. Dadurch bleiben dem Datennutzer inhaltlich wertvolle Informationen erhalten.

Zusätzlich zu den beschriebenen Maßnahmen wurden 7 Unternehmen (dies entspricht rund 0,05 % der Unternehmen aus der Stichprobe) aus den beiden relativ dünn besetzten Wirtschaftszweigen „523“ *Apotheken; Facheinzelhandel mit medizinischen, orthopädischen und kosmetischen Artikeln (in Verkaufsräumen)* und „5 241“ *Einzelhandel mit Textilien* durch weitere Anonymisierung geschützt. Ein ostdeutsches Unternehmen aus dem Wirtschaftszweig „525“ wurde aus dem Datensatz herausgenommen. Bei drei westdeutschen Unternehmen aus dem Wirtschaftszweig „525“ wurden die Merkmale Gesamtumsatz und Beschäftigte jeweils gemittelt (punktuelle Mikroaggregation). Bei drei ostdeutschen Unternehmen aus dem Wirtschaftszweig „5 241“ wurden die Ausprägungen im Merkmal Beschäftigte gemittelt.

Durch den Verzicht auf datenverändernde Anonymisierungsverfahren erhält der Datennutzer die metrischen Originalwerte. Einschränkungen ergeben sich daher lediglich durch die traditionellen Anonymisierungsmaßnahmen. Beispielsweise sind Regionalanalysen „nur“ nach Ost- und West-

17) Das sind die Unternehmen mit einem Mindestjahresumsatz von 250 000 Euro. – 18) Siehe Fußnote 17 und Vorgrimler, D., Dittrich, S., Lenz, R., Rosemann, M.: „Ein Scientific-Use-File der Umsatzsteuerstatistik 2000“, in *Wirtschaft und Statistik* 3/2005, 201 – 210.

deutschland möglich. Wenn ein Forscher an anderen Regionalanalysen interessiert ist, kann er über das Forschungsdatenzentrum verschiedene Zugangsmöglichkeiten dafür beantragen.

Eine Liste aller im Datensatz enthaltenen Merkmale findet sich in Anhang 2.¹⁹⁾

3.2.2 Scientific-Use-File für alle Unternehmen der Einzelhandelsstatistik

Mit den im letzten Projektabschnitt getesteten Verfahren der Zufallsüberlagerung ist es gelungen, einen weiteren Scientific-Use-File ohne Abschneidegrenze hinsichtlich der Anzahl an Beschäftigten in den Unternehmen zu erzeugen. Wie bei dem Scientific-Use-File für kleine Unternehmen beträgt der Mindestjahresumsatz 250 000 Euro.

Grundsätzlich wurden hier dieselben informationsreduzierenden Maßnahmen wie zuvor beschrieben eingesetzt. Zusätzlich wurde die Regionalinformation entfernt. Als datenveränderndes Verfahren wurde eine Variante aus der Verfahrensgruppe der Zufallsüberlagerungen angewendet und folgende Schritte dabei durchgeführt: Multiplikative Überlagerung der Einzelwerte nach dem Verfahren von Höhne mit den Parametern $f = 0,08$ und $s = 0,02$.²⁰⁾ Damit werden die Angaben des Betriebes entweder einheitlich erhöht oder gesenkt (mit Wahrscheinlichkeit von je 50 % und im Mittel um 8 %), wobei der genaue Faktor der Veränderung im Durchschnitt um 2 Prozentpunkte von den 8 % abweicht.

Da die bisher beschriebenen Maßnahmen nicht in allen Zellen einen ausreichenden Schutz garantierten, wurden folgende punktuellen Eingriffe vorgenommen: Mikroaggregation der drei Marktführer in den Wirtschaftszweigen „522“, „523“, „525“ und „5 247“ sowie Mikroaggregation der sechs Marktführer im Wirtschaftszweig „527“. Dabei wurden neben den Merkmalen Umsatz und Beschäftigte die mit diesen beiden stark korrelierten Merkmale (Korrelation $> 0,75$) ebenfalls mikroaggregiert. Dabei handelt es sich um die Merkmale Anfangs- und Endbestand an Handelsware, Sonstige betriebliche Erträge, Bezüge von Handelsware, Löhne und Gehälter, Sozialabgaben, Mieten und Pachten einschl. Kosten für Operate Leasing und Bezogene Leistungen und andere betriebliche Aufwendungen. Insgesamt sind 18 Merkmalsträger von den zusätzlichen Maßnahmen betroffen.

3.3 Umsatzsteuerstatistik

Im Rahmen der Umsatzbesteuerung werden von den Unternehmen Umsatzsteuer-Voranmeldungen bei den Finanzbehörden abgegeben.²¹⁾ Durch die Auswertung der monatlichen und vierteljährlichen Umsatzsteuer-Voranmeldungen ergänzt um Daten des Grundinformationsdienstes²²⁾ gewinnt die amtliche Statistik Informationen über die Entstehung der Umsatzsteuer,

19) Eine ausführliche Beschreibung des Scientific-Use-Files findet sich in Scheffler, M.: „Ein Scientific-Use-File der Einzelhandelsstatistik 1999“ in *Wirtschaft und Statistik* 3/2005, 197 – 200. – 20) Eine detaillierte Beschreibung des Verfahrens findet sich im Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten, siehe Fußnote 5. – 21) Neben den Umsatzsteuervoranmeldungen werden jährliche Umsatzsteuererklärungen abgegeben, die der endgültigen Steuerfestsetzung (Veranlagung) dienen. Die Voranmeldspflicht entfällt bei Unternehmen, die im vorangegangenen Jahr höchstens 511 Euro Umsatzsteuer zu entrichten hatten (sog. Jahreszahler). – 22) Im Grundinformationsdienst werden von der Finanzverwaltung Angaben zu den persönlichen Verhältnissen der Steuerpflichtigen gespeichert – z. B. Wohnort oder Gewerkekennzahl.

Das am häufigsten genutzte qualitative Merkmal der Umsatzsteuerstatistik ist die Wirtschaftszweigzuordnung. Maßgebend für die Zuordnung zu einem Wirtschaftszweig ist der Schwerpunkt der wirtschaftlichen Tätigkeit eines Unternehmens. Seit dem Berichtsjahr 2002 wird für die Umsatzsteuerstatistik die aktualisierte Ausgabe 2003 der Klassifikation der Wirtschaftszweige (WZ 2003) in der für die Steuerstatistiken gültigen Tiefengliederung zu Grunde gelegt (GKZ 2003). Für das Berichtsjahr 2000, für das der Scientific-Use-File vorliegt, ist die Ausgabe 1993 (WZ 93) relevant.²⁴⁾

Um die zur Weitergabe an die Wissenschaft notwendige faktische Anonymität der Merkmalsträger nach § 16 (6) BStatG zu gewährleisten, mussten Informationen aus dem Originalmaterial der Umsatzsteuerstatistik vergrößert, gelöscht und in wenigen Fällen verändert werden. Der Einsatz von datenverändernden Verfahren (in diesem Fall der Mikroaggregation) wurde hierbei auf ein Minimum beschränkt.

Die qualitativen Merkmale, zu denen wirtschaftliche und regionale Merkmale gehören, sind für den Erhalt der Vertraulichkeit besonders dann problematisch, wenn sie aufgrund einer tiefen Gliederung die Merkmalsträger in einer solchen Weise beschreiben, dass nur noch wenige Unternehmen dieser Beschreibung entsprechen (z. B. der einzige Bäcker einer Gemeinde).

Aus diesem Grunde sieht das Anonymisierungskonzept bei den qualitativen Merkmalen folgende Maßnahmen vor:

- Die Regionalangabe wird auf Ost/West vergrößert, wobei der Osten die neuen Bundesländer inklusive Berlin umfasst.
- Die Wirtschaftszweigzuordnung wird in unterschiedlicher Tiefengliederung abhängig von den Besetzungszahlen in die Daten aufgenommen. Dabei werden z. T. auch neue zusammengefasste Positionen gebildet.
- Aus dem Merkmal Beginn der Steuerpflicht wird das Merkmal Neugründung mit den Ausprägungen 1 = ja und 0 = nein gebildet. Bei Unternehmen mit mehr als 100 Mio. € Umsatz, wird das Merkmal generell auf 0 gesetzt.²⁵⁾
- Das Merkmal Rechtsform wird vergrößert (mit den Ausprägungen Personengesellschaften, Kapitalgesellschaften, Genossenschaften und sonstige Rechtsformen).

Bei den quantitativen Merkmalen (z. B. Umsatz) entstehen besonders bei extremen Ausprägungen datenschutzrechtliche Probleme. So ist z. B. das Unternehmen, das den höchsten Umsatz erzielt, in besonderer Weise gefährdet. Verschärft wird dieses Problem bei einer Kombination von quantitativen und qualitativen Merkmalen. Wird z. B. der Umsatz innerhalb eines Wirtschaftszweiges bzw. einer Branche betrachtet, ist der jeweilige Branchenmarktführer besonders gefährdet. Daher besteht das Anonymisierungskonzept bei den quantitativen Merkmalen aus einer zweistufigen Mikroaggregation. Die erste Stufe beinhaltet eine für jedes Merkmal getrennt

²⁴⁾ Ergebnisse zu allen 830 nachgewiesenen Wirtschaftszweigen können kostenlos über den Statistik-Shop des Statistischen Bundesamtes bezogen werden (www.ec.destatis.de). Siehe auch Dittrich, S.: „Umsätze und ihre Besteuerung 2002“ in *Wirtschaft und Statistik* 10/2004, S. 1195 – 1200. – ²⁵⁾ Von über 150.000 als Neugründungen gekennzeichneten Unternehmen haben 118 Unternehmen einen Umsatz von über 100 Mio. €. Bei diesen wird dieses Merkmal auf 0 gesetzt und damit die Information unterdrückt. Aus Plausibilitätsgründen dürfte die Informationsreduktion nicht besonders relevant sein, da es sich in der Mehrheit der Fälle um keine echten Neugründungen handeln wird. Zu „unechten“ Neugründungen kommt es bspw. bei Rechtsformänderungen oder Sitzverlagerungen.

durchgeführte Mikroaggregation für alle Unternehmen.²⁶⁾ Dies gilt wie erwähnt als die schwächste Form der Mikroaggregation.²⁷⁾ Da durch diese getroffenen Maßnahmen die regionalen Branchenmarktführer²⁸⁾ aber noch nicht genügend geschützt werden, wurde zusätzlich eine punktuelle Mikroaggregation durchgeführt. Dabei werden nur speziell die jeweiligen drei regionalen Marktführer einer Branche mikroaggregiert, wobei das Merkmal Lieferungen und Leistungen (Umsatz) das bestimmende Merkmal ist. Da in dem Scientific-Use-File zwischen 68 Wirtschaftszweigen und zwei Regionen unterschieden wird, sind insgesamt 408 Merkmalsträger von dieser zweiten Stufe betroffen (jeweils drei Unternehmen in den 136 regionalisierten Wirtschaftszweigen).

Je herausragender eine Ausprägung ist, desto stärker weicht der mikroaggregierte Wert vom Originalwert ab. Bei kleineren Unternehmen unterscheiden sich die Umsätze kaum. Bei einem Großunternehmen unterscheidet sich hingegen der Umsatz evtl. deutlich von dem Umsatz des nächst größten Unternehmens. Daher werden bei der Mikroaggregation die Merkmalsausprägungen der kleinen und mittleren Unternehmen nur minimal verändert, die der großen hingegen stärker. Je nach Merkmal sind zwischen 90 und 99 % der Ausprägungen trotz der Mikroaggregation unverändert in den Scientific-Use-File eingegangen. Nur sehr wenige weisen relative Abweichungen von mehr als 5 % zwischen mikroaggregiertem und originalem Wert auf. Betrachtet man jedoch lediglich die Unternehmen mit mehr als 500 Mio. € Umsatz, dann sind gut die Hälfte der Ausprägungen durch die Anonymisierung verändert worden, allerdings ca. 90 % innerhalb einer engen Schranke von 5 % relativer Abweichung. Da die Großunternehmen gefährdeter hinsichtlich einer Reidentifikation sind, erfüllt diese Art der Anonymisierung besonders die Forderung, Merkmalsträger hinsichtlich ihres Reidentifikationsrisikos zu anonymisieren.

Eine Liste aller im Datensatz enthaltenen Merkmale findet sich in Anhang 3.²⁹⁾

4 Ausblick

Der Anspruch, den die Initiatoren des Projektes FAWE an ihre Arbeit hatten, ging über die Schaffung eines konkreten Datenangebotes für bestimmte Erhebungen der statistischen Ämter hinaus. Wesentliche Leistung der durchgeführten Grundlagenforschung ist es, den Haltern von Mikrodaten eine ressourcenschonende Anonymisierung zu ermöglichen, die es ihnen erlaubt, ihre Datenbestände Forschern mit vertretbarem Aufwand zugänglich zu machen. Für die statistischen Ämter als wesentlichen Datenhaltern bedeutet dies konkret, dass sie die vom Bundesstatistikgesetz geforderte faktische Anonymität von Mikrodaten erzeugen können. Dazu wird hier ein Kompendium der Methoden der Anonymisierung und Hilfen für deren sinnvolle und effiziente Anwendung bereitgestellt. Kontakte zu anderen Datenhaltern im In- und Ausland haben bereits gezeigt, dass hieran ein erhebliches Interesse besteht. Das Kompendium, das mit dem kürzlich veröffentlichten Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten³⁰⁾ vorgelegt wird, liefert den

26) Zum Verfahren der Mikroaggregation siehe Fußnote 1. – 27) Zur Wirkung der Mikroaggregation bei der Umsatzsteuerstatistik siehe Lenz, R., Vorgrimler, D.: „Geheimhaltungsmethoden auf dem Prüfstand – eine Analyse anhand der Umsatzsteuerstatistik“ in *Wirtschaft und Statistik* 6/2004, S. 639 – 648. – 28) Mit den regionalen Branchenmarktführern sind die jeweils drei Unternehmen gemeint, die in den jeweiligen Branchen (abgegrenzt nach der Wirtschaftsklassifikation) innerhalb ihrer Region (Ost/West) die höchsten Umsätze erzielen. – 29) Einige Beispielsanalysen mit dem Scientific-Use-File finden sich in Vorgrimler, D., Dittrich, S., Lenz, R., Rosemann, M.: „Wissenschaftliche Analysen mit Hilfe der amtlichen Umsatzsteuerstatistik“, *Wirtschaftswissenschaftliches Studium*, Heft 10/2005, 585 – 590.

Ämtern und allen weiteren interessierten Datenanbietern Hilfen für eine sinnvolle und effiziente Anonymisierung. Es bietet die Grundlage für die Methodenauswahl, die Anwendung der Methoden und die Prüfung ihrer Wirkungsweise hinsichtlich künftiger Anonymisierungen.

Die statistischen Ämter haben mehrere Felder weiterer Datenangebote ins Auge gefasst. Zum einen sollen weitere Jahrgänge der dargestellten Erhebungen und auch andere Erhebungen anonymisiert werden. Zum anderen wird ein weiteres Forschungsprojekt „Wirtschaftsstatistische Paneldaten und faktische Anonymisierung“, das die Datenanbieter ab Januar 2006 unter Förderung des BMBF mit der Wissenschaft durchführen, den Aspekt der Anonymisierung von Paneldaten bearbeiten.

Anhang 1

Merkmale der Kostenstrukturerhebung im Verarbeitenden Gewerbe:

1. Wirtschaftsabteilung (WZ 93 auf Zweistellerebene)
2. Regionalbezug (Ost-West Klassifizierung)
3. Beschäftigtengrößenklasse
 - 07 = 20 – 49
 - 08 = 50 – 99
 - 11 = 100 – 249
 - 14 = 250 – 499
 - 17 = 500 – 999
 - 22 = 1000 und mehr
4. Teilzeitbeschäftigte
5. Teilzeitbeschäftigte umgerechnet in Vollzeiteinheiten
6. Tätige Personen insgesamt
7. Umsatz aus eigenen Erzeugnissen
8. Umsatz aus Handelsware
9. Gesamtumsatz (entspricht nicht der Summe aus 7. und 8.)
10. Anfangsbestand an unfertigen und fertigen Erzeugnissen aus eigener Produktion gemessen am Umsatz aus eigenen Erzeugnissen
11. Endbestand an unfertigen und fertigen Erzeugnissen aus eigener Produktion gemessen am Umsatz aus eigenen Erzeugnissen
12. Gesamtleistung/Bruttoproduktionswert
13. Anfangsbestand an Rohstoffen und sonstigen fremdbezogenen Vorprodukten, Hilfs- und Betriebsstoffen gemessen am Umsatz aus eigenen Erzeugnissen
14. Endbestand an Rohstoffen und sonstigen fremdbezogenen Vorprodukten, Hilfs- und Betriebsstoffen gemessen am Umsatz aus eigenen Erzeugnissen
15. Verbrauch an Rohstoffen
16. Energieverbrauch
17. Anfangsbestand an Handelsware gemessen am Umsatz aus Handelsware
18. Endbestand an Handelsware gemessen am Umsatz aus Handelsware
19. Einsatz an Handelsware
20. Bruttogehalts- und -lohnsumme

30) siehe Fußnote 5.

21. Gesetzliche Sozialkosten
22. Sonstige Sozialkosten
23. Kosten für Leiharbeitnehmer
24. Kosten für Lohnarbeiten
25. Kosten für Reparaturen
26. Mieten und Pachten
27. Sonstige Kosten
28. Fremdkapitalzinsen
29. Kosten insgesamt
30. Bruttowertschöpfung zu Faktorkosten
31. Nettowertschöpfung zu Faktorkosten
32. Gesamtaufwendungen für innerbetriebliche Forschung und Entwicklung
33. Anzahl der für Forschung und Entwicklung eingesetzten Lohn- und Gehaltsempfänger

Anhang 2

Merkmale der Einzelhandelsstatistik:

1. Wirtschaftszweig (WZ 93 auf Drei- bzw. Vierstellerebene)
2. Gesamtumsatz
3. Umsatzanteile in % aus
 - Großhandel
 - Einzelhandel, Reparatur von Gebrauchsgütern
 - Sonstigen Dienstleistungstätigkeiten
 - Herstellung, Verarbeitung, anderen industriellen Tätigkeiten oder aus Land- und Forstwirtschaft und Fischerei
4. Sonstige betriebliche Erträge
5. Einzelhandelsumsatz nach Absatzformen
 - In Verkaufsräumen
 - Aus Versandhandel
 - An Verkaufsständen und auf Märkten
 - Aus sonstigem Einzelhandel
6. Bestände an Handelsware jeweils am Anfang und Ende des Geschäftsjahres
7. Dito Roh-, Hilfs- und Betriebsstoffe
8. Dito selbsthergestellte und bearbeitete Halb- und Fertigerzeugnisse
9. Bezüge von Handelsware
10. Bezüge von Roh-, Hilfs- und Betriebsstoffen
11. Löhne und Gehälter
12. Sozialabgaben
13. Mieten und Pachten einschl. Kosten für Operate Leasing
14. Betriebliche Steuern und Abgaben
15. Bezogene Leistungen und andere betriebliche Aufwendungen
16. Bruttoinvestitionen in
 - Grundstücke
 - Bestehende Gebäude
 - Errichtung, Umbau und Erweiterung von Gebäuden
 - Maschinen, Einrichtungen und Fahrzeuge

17. Verkäufe von Sachanlagen
18. Wert der im Geschäftsjahr über Finanzierungsleasing erworbenen Sachanlagen
19. Zahl der rechtlich unselbständigen örtlichen Einheiten des Unternehmens am 31.12.
(in Kategorien)
20. Zahl der Beschäftigten am 30. 9.
Beschäftigte insgesamt
Darunter Lohn- und Gehaltsempfänger
Darunter Teilzeitbeschäftigte
21. Hochrechnungsfaktor
22. Punktuell mikroaggregiert (Dummy-Variable)

Anhang 3

Merkmale der Umsatzsteuerstatistik:

Regionalbezug (BBR-Schlüssel, sog. „Neuner-Kategorie“)
Wirtschaftszweig (WZ93)
Dauer der Steuerpflicht (typisiert)
Organschaft nach § „Abs. 2 Nr. 2 UStG (0= nein, 1 = ja)“
Rechtsform
Wertangaben in Euro
Lieferungen und Leistungen
Steuerpflichtige Lieferungen und Leistungen
Zu 16 %
Zu 7 %
Steuerfreie Lieferungen und Leistungen
Mit Vorsteuerabzug
Inneregemeinschaftlich Lief. und Leist.
Ohne Vorsteuerabzug
Umsatzsteuer vor Abzug der Vorsteuer
Für Lieferungen und Leistungen
Für innergemeinschaftliche Erwerbe
Abziehbare Vorsteuer
Für Lieferungen und Leistungen
Aus Rechnungen anderer Unternehmen
Einfuhrumsatzsteuer
Für innergemeinschaftliche Erwerbe
Vorauszahlungssoll
Nachrichtlich: innergemeinschaftliche Erwerbe
Vorjahreswerte in Euro
Lieferungen und Leistungen
Vorauszahlungssoll

Autoren- und Kontaktangaben:

Roland Sturm, Gruppe IVA, „Unternehmensregister, Koordinierung der Unternehmensstatistiken, Klassifikationen“, Statistisches Bundesamt, Gustav-Stresemann-Ring 11, 65180 Wiesbaden, E-Mail: roland.sturm@destatis.de

Dr. Rainer Lenz, FDZ des Statistischen Bundesamtes, Gustav-Stresemann-Ring 11, 65180 Wiesbaden, E-Mail: rainer.lenz@destatis.de

Qualitätssicherung des Faktors Interviewer für den Mikrozensus

Christin Schäfer, Hartmut Bömermann, Ricarda Nauenburg, Karsten Wenzel,
Klaus-Robert Müller, Gert G. Wagner,

Dieser Beitrag beschreibt eine neue Vorgehensweise zum Qualitätsmonitoring für den Mikrozensus. Die Absicherung speziell des Faktors Interviewer steht im Fokus des Interesses. Die durchgeführten Analysen unterscheiden sich konzeptuell von bisherigen Verfahrensweisen, da sie die Daten des Mikrozensus selbst nutzen, um Aussagen über etwaige Unkorrektheiten in der Durchführung des Mikrozensus aufzudecken. Hierbei werden neben Methoden der klassischen Statistik auch Methoden des Maschinellen Lernens eingesetzt. Das gesamte Verfahren kann prinzipiell auf der bestehenden Hardware des Statistischen Landesamtes durchgeführt werden, eine Implementierung kann an vorhandene Software ansetzen.

1 Einleitung und Problemstellung

Dieser Beitrag stellt die ersten Ergebnisse eines Kooperationsprojektes zwischen dem Deutschen Institut für Wirtschaftsforschung (DIW Berlin), dem Statistischen Landesamt Berlin und dem Fraunhofer Institut für Rechnerarchitektur und Softwaretechnik FIRST.IDA dar. Anstoß für die Zusammenarbeit war ein erfolgreiches Projekt zwischen dem DIW und FIRST.IDA; in diesem Projekt wurden fälschende Interviewer mittels einer Analyse der Daten des sozio-ökonomischen Panels (SOEP) identifiziert. Ziel des hier vorgestellten Projektes ist die Klärung der Frage, inwieweit mit datenanalytischen Methoden auch für den Mikrozensus eine Qualitätskontrolle des Faktors Interviewer durchgeführt werden kann.

In Zeiten einschneidender Einsparungen in den statistischen Ämtern, bei gleichzeitig steigenden Anforderungen an die Verfügbarkeit von Daten und einer erheblich intensiveren Nutzung der Mikrodaten durch die empirische Forschung stellt sich die Frage, wie die Qualität der Mikrozensus-erhebung gesichert bzw. weiter verbessert werden kann. Eine hohe Datenqualität ist beim Mikrozensus unabdingbar, da der Mikrozensus die bedeutendste Erhebung zu den Lebens- und Arbeitsbedingungen der Bevölkerung in Deutschland ist und auf seiner Grundlage weitreichende politische Entscheidungen getroffen werden (Statistische Ämter des Bundes und der Länder 2003; Statistisches Bundesamt, 2005), Quality Concept for Official Statistics: Encyclopedia of Statistical Sciences). Die meisten gegenwärtig angewendeten Standardmethoden zur Qualitätssicherung sind zeit- und personalintensiv. Beide Ressourcen sind sehr knapp; nur ihr gezielter Einsatz verspricht positive Effekte auf die Qualität des Mikrozensus. Datenanalytische Methoden erlauben hier ein zeit- und kostengünstiges effektives Vorgehen.

Unter dem Aspekt der Qualitätssicherung ist dem Faktor Interviewer erst in den letzten Jahren mehr Aufmerksamkeit geschenkt worden (Reuband, 1990; Schnell, 1991; Koch, 1995; Diekmann, 2002; Lynn, 2004; Schäfer et al., 2005). Nur ungern werden in diesem Bereich Probleme eingeräumt. Studien in der empirischen Sozialforschung haben jedoch gezeigt, dass Fälschun-

gen durch Interviewer überall gefunden wurden, wo auch nach ihnen gesucht wurde. Ferner zeigte sich, dass der Anteil der zumindest teilgefälschten Interviews erschreckend hoch liegt (Diekmann, 2002, berichtet in einem Beispiel von 20 % Fälschungen) und diese Fälschungen einen negativen Einfluss auf die Datenqualität haben, der nicht vernachlässigt werden kann (Schräpler und Wagner, 2005).

2 Daten und Datenzugang

Der Mikrozensus ist mit einem Auswahlsatz von 1 % die größte nationale Mehrthemenumfrage. In Berlin werden jährlich etwa 18 000 Haushalte mit 34 000 Personen befragt. Das vollständige Frageprogramm umfasste im Jahr 2004 ca. 160 Fragen. Für die Feldarbeit wurde in Berlin ein Stab von 170 Interviewern eingesetzt.

Die Bearbeitung des Projektes erfolgte im Rahmen eines Gastaufenthalts im Statistischen Landesamt Berlin. Für die Durchführung wurde die Infrastruktur des Forschungsdatenzentrums am Regionalen Standort Berlin genutzt.

3 Methodik

Im SOEP-Projekt kamen zur Erkennung von fraudulent Interviewern zwei Methoden zum Einsatz: Benfords Law und die Variability Method. Beide Methoden berechnen für jeden Interviewer eine Kenngröße. Nachfolgend wird die Wahrscheinlichkeit berechnet, den beobachteten Wert der Kenngröße auf der vorliegenden Stichprobe zu beobachten. Mittels dieser Wahrscheinlichkeit erfolgt ein Ranking der Interviewer. Interviewer, die einen besonders verdächtigen Cluster von Fragebögen abgeliefert haben, bilden den Anfang des Rankings. Ihre Überprüfung kann gezielt durchgeführt werden.

Im Folgenden sollen kurz die beiden verwendeten Verfahren dargestellt und ihre Einsatzmöglichkeit für eine Analyse des Mikrozensus diskutiert werden.

3.1 Benfords Law

Bei Benfords Law handelt es sich um ein empirisches Gesetz, das besagt, dass die Verteilung der ersten Ziffer aller Zahlen vieler Datensätze mit metrischen Größen einer logarithmischen Verteilung folgen, wobei die Ziffer 1 mit höchster Wahrscheinlichkeit, die Ziffer 9 mit niedrigster Wahrscheinlichkeit zu beobachten ist. Im Jahr 1995 lieferte Hill die theoretische Begründung für diese empirische Beobachtung.

Abweichungen der beobachteten von den theoretisch vorhergesagten Verteilungen liefern Hinweise auf mögliche Fälschungen, ein Verfahren, das beispielsweise auch in der Bilanzprüfung eingesetzt wird. Entscheidend für die Anwendbarkeit von Benfords Law ist, dass die Größen des Datensatzes metrisch mit positivem unbeschränktem Wertebereich sind.

Diese Bedingung ist für den Mikrozensus nicht erfüllt. Das Gros der Variablen ist nominal oder ordinal skaliert. Eine Beschränkung der Analyse auf die wenigen quasi-metrischen Größen reduziert den Datensatz so deutlich, dass davon ausgegangen werden kann, dass auch die in den Daten enthaltenen Informationen über die fälschenden Interviewer deutlich reduziert wurden und daher keine Schlüsse mehr möglich sind. Aus diesem Grund wurde in diesem Projekt auf die Anwendung von Benfords Law verzichtet. Die Analyse konzentrierte sich statt dessen auf die Variability Method.¹⁾

3.2 Variability Method

Die der Variability Method zugrunde liegende Annahme ist, dass die von Interviewern gefälschten Fragebögen in ihrer Gesamtheit weniger Variabilität der Antworten aufweisen als eine vergleichbare Zufallsstichprobe aus der Menge aller Fragebögen. Dies subsumiert die Annahmen, dass fälschende Interviewer zu wenig fehlende Werte produzieren, keine fehlerhaften Aussagen machen (da sie keine Fragen missverstehen und den Fragebogen und seine Struktur kennen) und bei Einschätzungsfragen extreme Aussagen vermeiden. Die produzierten Interviews verstecken sich im unauffälligen Mittel. Aus dieser Sicht betrachtet kann man von inneren Punkten sprechen. Ziel ist es nun, ein Kriterium zu finden, durch das die inneren Punkte zu Ausreißern transformiert werden, welche einfach erkannt, gesondert behandelt und gegebenenfalls aus den Daten entfernt werden können.

Die in der verwendeten Statistik modellierte Idee ist die folgende: Jedes Merkmal weist auf der Gesamtstichprobe eine bestimmte Varianz der Antworten auf. Ein Cluster von Fragebögen sollte diese Varianz für jedes Merkmal reproduzieren, wobei die Abweichungen um so kleiner werden sollten, je mehr Fragebögen betrachtet werden. Summiert man die beobachteten Varianzen auf, ergibt sich eine Größe, die die Variabilität des betrachteten Fragebogenclusters widerspiegelt. Fasst man die Fragebögen eines Interviewers zu einem Cluster zusammen, erreichen die verdächtigen Interviewer bei der auf diesem Cluster berechneten Statistik zu kleine Werte. Sie sind damit klar zu identifizieren.²⁾ Zwei Punkte sind zu beachten: zum einen ist es zu empfehlen, Merkmale nicht zu betrachten, die nur von wenigen Befragten beantwortet wurden und bei denen folglich die Mehrzahl der Einträge ein und denselben Wert aufweist. Abhängig von der Struktur des vorliegenden Datensatzes empfiehlt sich ein Schwellenwert zwischen 80 % und 99 %. Des Weiteren ist es wichtig, die Merkmale so zu normieren, dass der Beitrag jedes Merkmals zur Gesamtvarianz gleich groß ist. Diese Maßnahmen verhindern erstens, dass allein ein oder zwei sehr untypische Fragebögen die Varianz eines Fragebogenclusters über einen kritischen Wert heben und zweitens, dass die Gesamtvarianz von den Varianzbeiträgen weniger quasi-metrischer Merkmale, beispielsweise Fragen nach dem Einkommen, dominiert wird.

1) Es ist Aufgabe späterer Analysen, den Einsatz eines empirischen Benford Tests zu erproben, bei dem sich die Verteilung der ersten Ziffer nicht nach Benfords Law berechnet, sondern nach der empirisch auf dem gesamten Datensatz beobachteten Verteilung. Dies kann von großem Wert bei den Mikrozensusdaten sein, da hier beispielsweise missing values häufig mit 8, 9 oder 99 kodiert werden. Dies führt zu einer extrem von Benfords Law abweichenden Häufigkeit für diese Ziffern. – 2) Die Analyse der SOEP-Daten hat gezeigt, dass dieser Test bereits Auffälligkeiten ans Licht bringen kann, wenn es sich nur bei einem Teil der Fragebögen eines Interviewers um Fälschungen handelt. Dieses Szenario ist sehr realitätsnah. Die Bereitschaft, ein Interview zu fabrizieren oder zumindest Teile eines Interviews ohne Befragung auszufüllen, wächst mit dem Zeitdruck, der am Ende eines Befragungszeitraums eintritt. In Einzelfällen erlaubte die Analyse sogar, innerhalb eines suspekten Fragebogenclusters die „wahren“ Interviews von den gefälschten zu unterscheiden.

Die Variabilität eines Clusters, d. h. aller von einem Interviewer durchgeführten Interviews, ergibt sich zusammengefasst aus der folgenden Formel:

$$T = \sum_{i=1}^{nM} \sum_{j=1}^{nQ} (x(i, j) - \bar{x}(i))^2$$

wobei $x(i, j)$ den Wert des j-ten Interviews bei der i-ten Frage und $\bar{x}(i)$ den Mittelwert der i-ten Frage gemittelt über alle Interviews bezeichnet. Die Summationen laufen einmal bis nQ , der Anzahl aller von einem Interviewer durchgeführten Interviews, und nM , der Anzahl aller in die Varianzberechnung einfließenden Variablen.

Beim Mikrozensus tritt nun folgendes zusätzliche Problem auf: das Kodierungsschema orientiert sich an den Möglichkeiten von Großrechnern. Insbesondere wird jedes Merkmal mit so wenig Stellen wie möglich kodiert, um Speicherplatz zu sparen. Dabei gilt die Regel, dass fehlende Werte als größtmöglicher Wert kodiert werden. Wird eine Frage nur mit einer Stelle kodiert, so bedeutet die „9“ einen fehlenden Wert. Weist das Merkmal allerdings 10 Ausprägungen auf, müssen zwei Stellen zur Kodierung verwendet werden und Missings werden mit „99“ kodiert. Dies führt in der obigen Statistik zu erheblichen Verzerrungen. Da jeder Beitrag quadratisch in die Gesamtvarianz eingeht, kann die Gesamtvarianz zu einem Großteil aus den Varianzbeiträgen der fehlenden Werte gebildet werden.

Um diesem Kodierungsbias entgegenzutreten, ist es nun möglich, individuell für jedes Merkmal die Kodierung der fehlenden Werte an die Kodierung der übrigen Einträge „heranzuziehen“. Hat ein Merkmal 12 Ausprägungen, die mit 1 bis 12 kodiert sind, so weist man den fehlenden Werten die 13 zu. Eine andere Möglichkeit ist, fehlende Werte generell mit -1 zu kodieren.

Das Kodierungsschema des Mikrozensus weist weitere Besonderheiten auf: beispielsweise werden n Merkmalsausprägungen nicht zwingend mit den Ziffern 1 bis n kodiert. Die kodierten Werte weisen Lücken auf, so dass auch hier einzelne Merkmalsausprägungen einen höheren Beitrag zur Varianz leisten können als andere. (Dies gilt natürlich bei einer varianzbasierten Statistik immer. Je weiter entfernt vom Mittelwert, desto höher der Beitrag zur Varianz. Allerdings verzerren die Lücken die Mittelwertschätzung und des Weiteren dann die Varianzschätzung in nicht gewünschter Weise.) Dies wiederum ließe sich durch ein Zusammenschieben des Kodierungsschemas, also das Beseitigen der Kodierungslücken, umgehen. Jedoch sind auch andere Vorgehensweisen vorstellbar und ebenfalls sinnvoll.

Es stellt sich die Frage, welche der dargestellten Möglichkeiten gewählt werden soll. Für die Anwendung jeder der obigen Datentransformationen lassen sich gute Argumente finden. Allerdings wird man keine Vorgehensweise finden, die aus jeder Sicht optimal ist.

Fest steht jedoch, dass der Wert der verwendeten Teststatistik T von der verwendeten Kodierung abhängt. Es handelt sich hier um die größte Schwachstelle der Statistik. Folgendes ist anzumerken: für Analysen, die an inhaltlichen Aussagen interessiert sind, ist die Berechnung des Mittelwerts eines nominal skalierten Merkmals – vorsichtig formuliert – ein Fehler. Das gilt natürlich auch für die Berechnung der Varianz eines solchen Merkmals. Weder Mittelwert noch Vari-

anz haben irgendeine inhaltliche Relevanz. Der Fokus unserer Analyse liegt nicht auf einer inhaltlichen Fragestellung, sondern es handelt sich um eine Mustererkennungsanwendung. Es gilt zu ermitteln, inwieweit ein Fragebogencluster die Muster der gesamten Umfrage reproduziert. Insbesondere sind wir daran interessiert, fakende Interviewer zu erkennen, die sich – so lautet unsere Hypothese – durch geringe Varianz ihrer Antworten auszeichnen. Aus diesem Blickwinkel ist die Berechnung des Mittelwerts und der Abweichungen vom Mittelwert auch für ein nominales Merkmal sinnvoll.

Da der Wert der Teststatistik von der Kodierung der Daten abhängt, ist zu klären, inwieweit diese unterschiedlichen Werte auch zu unterschiedlichen Ergebnissen in der Einschätzung der Qualität eines Interviewers führen. Eine solche Untersuchung ist komplex, sofern sie überhaupt mit wissenschaftlicher Rigorosität durchführbar ist.

Wir umgehen diese Untersuchung, indem wir eine Repräsentation der Daten verwenden, die kodierungsinvariant ist.

Die Idee ist einfach: jedes Merkmal wird durch so viele binäre Dummyvariablen repräsentiert wie Merkmalsausprägungen vorliegen. Dabei wird diejenige binäre Variable auf 1 gesetzt, die die im Interview vorliegende Ausprägung kodiert. Fehlende Werte werden bei diesem Vorgehen „normalen“ Ausprägungen gleichgestellt. Sämtliche oben angesprochene durch die Kodierung der Daten induzierte Probleme sind dadurch gelöst. Auch eine Normierung der einzelnen Merkmale im Hinblick auf den vom Merkmal maximal zu leistenden Beitrag zur Gesamtvarianz kann häufig entfallen. Der Nachteil dieser Vorgehensweise ist das extreme Aufblähen des Datensatzes. Für den Mikrozensus Berlin konnte der Datensatz noch auf einem PC mit sehr hoher Speicherkapazität bearbeitet werden. Bei größerem Stichprobenumfang kann sich hier ein Problem ergeben.

In der obigen Gleichung zur Berechnung der Teststatistik verändert sich durch die Datentransformation lediglich die Summation, die nun nicht mehr über die Anzahl der Merkmale, sondern über die Anzahl nB aller Binärvariablen läuft:

$$T = \sum_{i=1}^{nB} \sum_{j=1}^{nQ} (x(i, j) - \bar{x}(i))^2$$

Es ist nun möglich, für jeden Interviewer den Wert der Teststatistik für das von ihm abgelieferte Fragebogencluster zu berechnen.

Die Einordnung des Wertes der Teststatistik – die Zuweisung der Plausibilität – erfolgt mittels Schätzung der Verteilung der Teststatistik auf den gesamten Umfragedaten durch einen Resampling-Ansatz. Hat ein Interviewer nQ Fragebögen abgeliefert, so werden aus der Gesamtheit aller Fragebögen sehr oft nQ Fragebögen zufällig gezogen und der Wert der Teststatistik, den diese Zufallsstichprobe liefert, berechnet. Der Vergleich mit dieser geschätzten Verteilungsfunktion liefert für jeden Interviewer einen Plausibilitätswert: das Quantil der geschätzten Verteilung, an der der Wert der Teststatistik des Interviewers liegt.

Die Interviewer werden nun anhand dieses Plausibilitätswertes geordnet. Kleine Werte weisen auf eine kritische, d. h. zu kleine Variabilität in den Fragebogenclustern hin. Das bedeutet nicht, dass jeder Interviewer am Anfang dieses Rankings ein Fälscher ist. Die zu geringe Variabilität kann auch durch eine homogene Befragungsgruppe verursacht sein, etwa durch einen sehr hohen Anteil an Interviews in einem Altersruhesitz. Je größer die Anzahl der abgelieferten Fragebögen ist, desto unwahrscheinlicher wird das Auftreten von extrem homogenen Befragungsgruppen. Generell gilt, dass die Qualität und Stärke der Aussage mit der Größe des Fragebogenclusters zunimmt. Über einen Interviewer, der nur einen Fragebogen abgegeben hat, kann nur wenig ausgesagt werden.

Ist bekannt, dass Interviewer aufgrund der Stichprobenziehung und des Versuchsplans sehr homogene Befragungsgruppen interviewen müssen, können beim Resampling solche Randbedingungen kontrolliert werden. Allerdings stellt dies einen großen Rechenaufwand dar und bisweilen kann es sein, dass die durch die Einschränkung der Randbedingungen verringerte Grundgesamtheit, aus der gezogen werden kann, zu klein wird.

Im Allgemeinen ist dieser Aufwand nicht nötig. Die Fachgruppen, die sich mit der Überprüfung der verdächtigen Interviewer auseinandersetzen, werden solche homogenen Randbedingungen sofort identifizieren.

4 Ergebnisse

Bisher konnte das so adaptierte Verfahren noch nicht auf die Mikrozensusdaten angewandt werden. Zur Zeit ist die eindeutige Zuordnung eines Interviewers zu den von ihm durchgeführten Interviews im Datenmaterial für die Analyse aus vielerlei Gründen nicht möglich. Beispielsweise teilen sich mehrere Interviewer einen Laptop und in den Daten ist nur noch der Laptop zu identifizieren. Auch werden Telefoninterviews, die von Mitarbeitern des Statistischen Landesamts durchgeführt wurden, zum Teil den Interviewern zugeordnet, aus deren Stichprobe der Interviewte stammt.

Mit der Erhebung 2006 wird eine Variable eingeführt, die diese Zuordnung erlaubt. Damit wird dann das Testen der vorgestellten Methode auch für den Mikrozensus möglich.

Erste Voruntersuchungen haben ergeben, dass noch weitere Adaptionen nötig sein werden. Insbesondere ist zu klären, inwieweit der Befragungsmodus eine Rolle spielt. Diese Fragen sind Gegenstand der aktuellen Forschung.

5 Diskussion

Dieser Artikel stellt ein Verfahren zur Identifizierung von fälschenden Interviewern einzig anhand einer Datenanalyse des Gesamtdatensatzes Mikrozensus dar. Das Verfahren ist schnell durchzuführen, kosten- und zeitsparend und gibt der Fachgruppe eine Orientierung zum gezielten Einsatz von Überprüfungs- und Qualitätssicherungsmaßnahmen.

Ist dieser Beitrag auch eine Anleitung zum perfekten Fälschen von Fragebögen? Wird ein geschickter Faker von nun an gezielt fehlende Werte produzieren, extreme Antworten generieren und seltene Lebenssituationen in seinen Fragebögen abbilden? Sicherlich können solche Tricks die Entdeckungswahrscheinlichkeit durch das Verfahren verringern. Allerdings ist das richtige Maß für die zu produzierende Variabilität unbekannt. Sie hängt von der Varianz der gesamten Umfrage ab. Es ist daher auch weiterhin schwierig, konsistente valide Fälschungen zu produzieren und dabei noch geschickt ausreichend Varianz zu erzeugen. Erfolgt zusätzlich noch eine Bewertung der Fragebögen mittels einer empirischen Version von Benfords Laws, so ist geschicktes Fälschen nahezu unmöglich. Der kluge Interviewer wird den Aufwand scheuen und einfach seiner Aufgabe nachkommen: Interviews führen.

Literatur

Diekmann, A. (2002). Diagnose von Fehlerquellen und methodische Qualität in der sozialwissenschaftlichen Forschung. Manuskript 06/2002, Institut für Technikfolgen-Abschätzung (ITA), Wien.

Hill, T. P. (1995). A Statistical Derivation for the Significant-Digit Law. *Statistical Science*, 10: 354 – 362.

Koch, A. (1995). Gefälschte Interviews: Ergebnisse der Interviewerkontrolle beim ALLBUS 1994. *ZUMA Nachrichten*, 36: 89 – 105.

Lynn, Peter (2004), Editorial: Measuring and communicating survey quality. *Journal of the Royal Statistical Society A*, 167 (4): 575 – 578.

Marshall, Eliot (2000): How Prevalent Is Fraud? That's a Million-Dollar Question. *Science magazine*, 290: 1662 – 1663.

Reuband, K.-H. (1990). Interviews, die keine sind – 'Erfolge' und 'Misserfolge' beim Fälschen von Interviews. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 4: 706 – 733.

Schäfer, C., J.-P. Schräpler, K.-R. Müller, G. G. Wagner (2005). Automatic Identification of Faked and Fraudulent Interviews in the German SOEP, *Schmollers Jahrbuch*, Duncker & Humblot, Berlin, 125: 183 – 193.

Schnell, R. (1991). Der Einfluss gefälschter Interviews auf Survey-Ergebnisse. *Zeitschrift für Soziologie*, 20(1): 25 – 35.

Schräpler, J.-P., G. G. Wagner (2005). Characteristics and Impact of Faked Interviews in Surveys-Analysis of genuine fakes in the raw data of SOEP. *Allgemeines Statistisches Archiv*, 89 (1): 7 – 20.

Statistische Ämter des Bundes und der Länder (2003): Qualitätsstandards in der amtlichen Statistik (URL: [//www.destatis.de/allg/d/ueber/q_stand.htm](http://www.destatis.de/allg/d/ueber/q_stand.htm)).

Statistisches Bundesamt (2005): Qualitätsbericht Mikrozensus (URL: [//www.destatis.de/download/qualitaetsberichte/qualitaetsbericht_mikrozensus.pdf](http://www.destatis.de/download/qualitaetsberichte/qualitaetsbericht_mikrozensus.pdf))

Quality Concept for Official Statistics: Encyclopedia of Statistical Sciences. John Wiley & Sons, New York, Update Vol. 3: 621 – 629 (IMF – Dissemination Standards Bulletin Board, URL: [//dsbb.imf.org/vgn/images/pdfs/Encyc.pdf](http://dsbb.imf.org/vgn/images/pdfs/Encyc.pdf)).

Autoren- und Kontaktangaben:

Christin Schäfer, Fraunhofer Institut FIRST.IDA, Kekuléstr. 7, 12435 Berlin,

E-Mail: christin.schaefer@first.fraunhofer.de

Hartmut Bömermann, Statistisches Landesamt Berlin, Alt-Friedrichsfelde 60, 10315 Berlin,

E-Mail: H.Boemermann@statistik-berlin.de

Ricarda Nauenburg, Statistisches Landesamt Berlin, Alt-Friedrichsfelde 60, 10315 Berlin,

E-Mail: R.Nauenburg@statistik-berlin.de

Karsten Wenzel, Statistisches Landesamt Berlin, Alt-Friedrichsfelde 60, 10315 Berlin,

E-Mail: K.Wenzel@statistik-berlin.de

Klaus-Robert Müller, Fraunhofer Institut FIRST.IDA, Kekuléstr. 7, 12435 Berlin,

E-Mail: klaus@first.fraunhofer.de

Gert G. Wagner, Deutsches Institut für Wirtschaftsforschung, Königin-Luise-Straße 5,

14195 Berlin, E-Mail: gwagner@diw.de

Datenangebot

(Stand: 15. Februar 2006)

Datenangebot

Die Mikrodaten können über die mit ● markierten Zugangswege bereits genutzt werden.
Mit ○ markierte Zugangswege werden derzeit vorbereitet und sind in Kürze ebenfalls nutzbar.*)

Statistik	Jahr	Public-Use-File	CAMPUS-File	Scientific-Use-File	On-Site-Nutzung in den Stat. Ämtern
-----------	------	-----------------	-------------	---------------------	-------------------------------------

Sozialstatistiken

Mikrozensus	2004	–	–	○	●
	2003	–	–	●	●
	2002	–	–	●	●
	2001	–	–	●	●
	2000	–	–	●	●
	1999	–	–	●	●
	1998	–	●	●	●
	1997	–	–	●	●
	1996	–	–	●	●
	1995	–	–	●	●
	1994	–	–	–	●
	1993	–	–	●	●
	1992	–	–	–	●
	1991	–	–	●	●
	1990	–	–	–	●
	1989	–	–	●	●
	1988	–	–	–	●
	1987	–	–	–	●
1986	–	–	–	●	
1985	–	–	–	●	
Zeitbudgeterhebung	2001/2002	–	–	●	●
	1991/1992V ¹⁾	–	–	●	●
	1991/1992	●	–	●	●
Europäisches Haushaltspanel	1996	–	–	●	●
	1995	–	–	●	●
	1994	–	–	●	●
Statistik der Eheschließungen	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
Statistik der Geburten	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●

*) Ein regelmäßig aktualisiertes Datenangebot finden Sie unter: www.forschungsdatenzentrum.de

1) V = Vergleichsdatensatz zu 2001/2002, um vergleichende Auswertungen zu ermöglichen.

Statistik	Jahr	Public-Use-File	CAMPUS-File	Scientific-Use-File	On-Site-Nutzung in den Stat. Ämtern
Noch: Sozialstatistiken					
Statistik der Sterbefälle	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
Wanderungsstatistik	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
Einbürgerungsstatistik ²⁾	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
Statistik der Studierenden – Wintersemester	2004/2005	–	–	–	○
	2003/2004	–	–	–	●
	2002/2003	–	–	–	●
	2001/2002	–	–	–	●
	2000/2001	–	–	–	●
	1999/2000	–	–	–	●
	1998/1999	–	–	–	●
	1997/1998	–	–	–	●
	1996/1997	–	–	–	●
	1995/1996	–	–	–	●
Statistik der Studierenden – Sommersemester	2005	–	–	–	○
	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
	1999	–	–	–	●
	1998	–	–	–	●
	1997	–	–	–	●
1996	–	–	–	●	
Statistik der Prüfungen – Wintersemester	2004/2005	–	–	–	○
	2003/2004	–	–	–	●
	2002/2003	–	–	–	●
	2001/2002	–	–	–	●
	2000/2001	–	–	–	●
	1999/2000	–	–	–	●
	1998/1999	–	–	–	●
	1997/1998	–	–	–	●
	1996/1997	–	–	–	●
1995/1996	–	–	–	●	

2) Ohne Hamburg und Schleswig-Holstein.

Statistik	Jahr	Public-Use-File	CAMPUS-File	Scientific-Use-File	On-Site-Nutzung in den Stat. Ämtern
Noch: Sozialstatistiken					
Statistik der Prüfungen – Sommersemester	2005	–	–	–	○
	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
	1999	–	–	–	●
	1998	–	–	–	●
	1997	–	–	–	●
	1996	–	–	–	●
Personal- und Stellenstatistik	2005	–	–	–	○
	2004	–	–	–	○
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
	1999	–	–	–	●
	1998	–	–	–	●
Statistik der Habilitationen	2005	–	–	–	○
	2004	–	–	–	○
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
	1999	–	–	–	●
	1998	–	–	–	●
Hochschulfinanzstatistik	2005	–	–	–	○
	2004	–	–	–	○
	2003	–	–	–	○
	2002	–	–	–	○
	2001	–	–	–	○
	2000	–	–	–	○
	1999	–	–	–	○
	1998	–	–	–	○
Europäische Erhebung zur beruflichen Weiterbildung (CVTS)	2000	–	–	●	●
Todesursachenstatistik	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
Krankenhausstatistik, Teil II: Diagnosen	2003	–	–	●	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●

Statistik	Jahr	Public-Use-File	CAMPUS-File	Scientific-Use-File	On-Site-Nutzung in den Stat. Ämtern
-----------	------	-----------------	-------------	---------------------	-------------------------------------

Noch: Sozialstatistiken

Statistik der Empfänger/-innen von laufender Hilfe zum Lebensunterhalt am Jahresende (Sozialhilfestatistik)	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	–
	2000	●	–	–	–
	1999	●	–	–	–
	1998	●	●	–	–
Statistik der Jugendhilfe, Teil I ^{3) 4)}	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
	1995	–	–	–	●
Statistik der Jugendhilfe, Teil II ⁴⁾	2004	–	–	–	●
	2000	–	–	–	●
	1996	–	–	–	●
Statistik der Jugendhilfe, Teil III ⁴⁾	2002	–	–	–	●
	1998	–	–	–	●
Statistik der Jugendhilfe, Teil IV ⁴⁾	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
	1995	–	–	–	●

Wirtschaftsstatistiken

Monatsbericht für Betriebe im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
	1999	–	–	–	●
	1998	–	–	–	●
	1997	–	–	–	●
	1996	–	–	–	●
	1995	–	–	–	●

3) Teil I – Bogen 4.3 der Statistik der Jugendhilfe ist noch nicht verfügbar. – 4) Es können – insbesondere für ältere Erhebungsjahre – nicht mehr für alle Bundesländer Mikrodaten zur Verfügung gestellt werden.

Statistik	Jahr	Public-Use-File	CAMPUS-File	Scientific-Use-File	On-Site-Nutzung in den Stat. Ämtern
Noch: Wirtschaftsstatistiken					
Vierteljährliche Produktionserhebung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
	1999	–	–	–	●
	1998	–	–	–	●
	1997	–	–	–	●
	1996	–	–	–	●
	1995	–	–	–	●
Erhebung für industrielle Kleinbetriebe im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	2002	–	–	–	○
	2001	–	–	–	○
	2000	–	–	–	○
	1999	–	–	–	○
	1998	–	–	–	○
	1997	–	–	–	○
	1996	–	–	–	○
	1995	–	–	–	○
Jahresbericht für Unternehmen im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	2004	–	–	–	○
	2003	–	–	–	○
	2002	–	–	–	○
	2001	–	–	–	○
	2000	–	–	–	○
	1999	–	–	–	○
	1998	–	–	–	○
	1997	–	–	–	○
	1996	–	–	–	○
1995	–	–	–	○	
Investitionserhebung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	2004	–	–	–	○
	2003	–	–	–	○
	2002	–	–	–	○
	2001	–	–	–	○
	2000	–	–	–	○
	1999	–	–	–	○
	1998	–	–	–	○
	1997	–	–	–	○
	1996	–	–	–	○
	1995	–	–	–	○
Jahreserhebung im Einzelhandel	2001	–	–	–	●
	2000	–	–	–	●
	1999	–	–	●	●

Statistik	Jahr	Public-Use-File	CAMPUS-File	Scientific-Use-File	On-Site-Nutzung in den Stat. Ämtern
Noch: Wirtschaftsstatistiken					
Monatserhebung im Tourismus	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
	1999	–	–	–	●
	1998	–	–	–	●
	1997	–	–	–	●
	1996	–	–	–	●
	1995	–	–	–	●
	1994	–	–	–	●
	1993	–	–	–	●
	Dienstleistungsstatistik auf Grundlage der Strukturverordnung der EU	2003	–	–	–
2002		–	–	–	●
2001		–	–	–	●
2000		–	–	–	●
Gehalts- und Lohnstrukturerhebung im Produzierenden Gewerbe und im Dienstleistungsbereich	2001	–	–	○	●
	1995	–	–	–	●
	1992	–	–	–	● ⁵⁾
	1990	–	–	–	● ⁶⁾
EU-Erhebung über Arbeitskosten im Produzierenden Gewerbe und im Dienstleistungsbereich	2000	–	–	–	●
	1996	–	–	–	○
	1992	–	–	–	○
Kostenstrukturerhebung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	1999	–	● ⁷⁾	● ⁷⁾	–
Statistik der Baufertigstellungen	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
Statistik der Baugenehmigungen	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
Gebäude- und Wohnungsstichprobe	1993	–	–	–	●

5) Neue Bundesländer, ohne Mecklenburg-Vorpommern. – 6) Alte Bundesländer, ohne Bremen, Hamburg, Hessen und Nordrhein-Westfalen. – 7) KSE in kleinen und mittleren Unternehmen.

Statistik	Jahr	Public-Use-File	CAMPUS-File	Scientific-Use-File	On-Site-Nutzung in den Stat. Ämtern
Finanz- und Steuerstatistiken					
Gewerbesteuerstatistik	1998	–	–	–	●
Körperschaftsteuerstatistik	1998	–	–	–	●
	1995	–	–	–	●
	1992	–	–	–	●
Lohn- und Einkommensteuerstatistik	1998	–	–	●	●
	1995	–	–	–	●
Umsatzsteuerstatistik	2001	–	–	–	●
	2000	–	–	●	–
Erbschaft- und Schenkungsteuerstatistik	2002	–	–	●	–
Jährliche Schulden der Gemeinden und Gemeindeverbände	2003	–	–	–	○
	2002	–	–	–	○
	2001	–	–	–	○
	2000	–	–	–	○
	1999	–	–	–	○
	1998	–	–	–	○
	1997	–	–	–	○
	1996	–	–	–	○
	1995	–	–	–	○
Vierteljährliche Kassenstatistik der Gemeinden und Gemeindeverbände	2004	–	–	–	○
	2003	–	–	–	○
Rechtspflegestatistiken⁸⁾					
Strafverfolgungstatistik	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
	1999	–	–	–	●
	1998	–	–	–	●
	1997	–	–	–	●
	1996	–	–	–	●
	1995	–	–	–	●
Strafvollzugsstatistik	2004	–	–	–	●
	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	●
	2000	–	–	–	●
	1999	–	–	–	●
	1998	–	–	–	●
	1997	–	–	–	●
1996	–	–	–	●	
1995	–	–	–	●	

8) Die Daten der Rechtspflegestatistiken stehen über die kontrollierte Datenfernverarbeitung zur Verfügung.

Statistik	Jahr	Public-Use-File	CAMPUS-File	Scientific-Use-File	On-Site-Nutzung in den Stat. Ämtern
Noch: Rechtspflegestatistiken⁸⁾					
Bewährungshilfestatistik	2004	–	–	–	○
	2003	–	–	–	○
	2002	–	–	–	○
	2001	–	–	–	○
	2000	–	–	–	○
	1999	–	–	–	○
	1998	–	–	–	○
	1997	–	–	–	○
	1996	–	–	–	○
	1995	–	–	–	○
Agrar- und Umweltstatistiken					
Statistik über die öffentliche Wasserversorgung	2001	–	–	–	●
	1998	–	–	–	●
Statistik über die öffentliche Abwasserbehandlung	2001	–	–	–	●
	1998	–	–	–	●
Statistik über die öffentliche Abwasserbeseitigung	2001	–	–	–	●
	1998	–	–	–	●
Statistik über die Wasserversorgung und Abwasserbeseitigung bei Wärmekraftwerken für die öffentliche Versorgung	2001	–	–	–	●
	1998	–	–	–	○
Statistik über die Wasserversorgung und Abwasserbeseitigung im Bereich Verarbeitendes Gewerbe, Bergbau und Gewinnung von Steinen und Erden	2001	–	–	–	●
	1998	–	–	–	○
Statistik über die Wassereigenversorgung und -entsorgung privater Haushalte	2001	–	–	–	●
	1998	–	–	–	●
Landwirtschaftszählung – Haupterhebung	1999	–	–	–	●
Agrarstrukturerhebung mit Bodennutzung und Viehzählung	2003	–	–	–	●
	2001	–	–	–	●
	1999	–	–	–	●
Integrierte Erhebung über Bodennutzung und Viehbestände	2003	–	–	–	●
	2002	–	–	–	●
	2001	–	–	–	○
	2000	–	–	–	●
	1999	–	–	–	●