

Falk, Martin; Steiner, Viktor

**Research Report**

Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik im Bereich Erwerbstätigkeit und Einkommen

ZEW Gutachten/Forschungsberichte

**Provided in Cooperation with:**

ZEW - Leibniz Centre for European Economic Research

*Suggested Citation:* Falk, Martin; Steiner, Viktor (2000) : Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik im Bereich Erwerbstätigkeit und Einkommen, ZEW Gutachten/Forschungsberichte, Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim

This Version is available at:

<https://hdl.handle.net/10419/110481>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# **Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik im Bereich Erwerbstätigkeit und Einkommen**

**Martin Falk** <sup>a)</sup>  
**Viktor Steiner** <sup>a), b)</sup>

**Zentrum für Europäische Wirtschaftsforschung (ZEW) Mannheim** <sup>a)</sup>  
**Johann-Wolfgang-Goethe-Universität Frankfurt/Main** <sup>b)</sup>

**Mannheim, im Juli 2000**

Kontaktperson: Dr. Viktor Steiner  
ZEW  
L 7, 1  
D-68161 Mannheim  
Tel.: 49 621 1235 151  
Fax.: 49 621 1235 225  
Email: [steiner@zew.de](mailto:steiner@zew.de)

## **Problemstellung und Zusammenfassung**

Diese für die Kommission zur „Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik“ verfasste Expertise gibt einen Überblick über die Datenlage im Bereich Erwerbstätigkeit und Einkommen in Deutschland. Sie stellt die Defizite der aktuellen Datenlage aus Sicht der empirischen Wirtschaftsforschung dar und gibt Empfehlungen zur Verbesserung dieser Situation. Die Darstellung der Datenlage berücksichtigt die Amtliche Statistik auf dem Gebiet der Bevölkerungs- und Arbeitslosenstatistiken, Lohn-, Gehalts- und Arbeitskostenstatistiken sowie Einkommens- und Sozialhilfestatistiken. Darüber hinaus werden auch nicht-amtliche Befragungsdaten mit in Betracht gezogen. Angesichts der Fülle von nicht-amtlichen Befragungsdaten, die von Forschungsinstituten, Ministerien, Unternehmensberatungen und Verbänden erhoben werden, werden nur Wiederholungsbefragungen berücksichtigt. Zudem müssen die Datensätze repräsentativ für eine größere Population der deutschen Bevölkerung bzw. der Gesamtheit aller Betriebe oder Unternehmen sein und sich auf das gesamte Bundesgebiet beziehen. Einzeldaten, die sich nur auf ein Bundesland beziehen, wie z. B. der Arbeitsmarktmonitor Sachsen-Anhalt oder das Hannoveraner Firmenpanel, werden hier nur ergänzend in einem Anhang berücksichtigt. Entsprechend der Problemstellung der vorliegenden Expertise stellt aus Sicht der empirischen Wirtschaftsforschung die Eignung und Verfügbarkeit der einzelnen Datenbasen für wissenschaftliche Zwecke das zentrale Kriterium dar. Dabei unterscheiden wir vier Gruppen von Datensätzen: (a) für die Wissenschaft frei verfügbare Mikrodaten, (b) eingeschränkt verfügbare Mikrodaten, (c) nicht verfügbare Mikrodaten und (d) aggregierte Daten. Das Schwergewicht der Expertise liegt eindeutig auf den Mikrodaten.

Die Bestandsaufnahme und Bewertung der Datenlage aus Sicht der empirischen Wirtschaftsforschung hat ergeben:

1. Die internationale empirische Forschung im Bereich Erwerbstätigkeit und Einkommen basiert ganz überwiegend auf Mikrodaten, aggregierte Daten haben in den letzten Jahrzehnten stark an Bedeutung verloren. Mikrodaten erlauben eine sachgerechte Abgrenzung der Untersuchungseinheiten, bilden die Heterogenität zwischen den Untersuchungseinheiten ab und haben daher einen ungleich höheren Informationsgehalt als aggregierte Daten. Darüber hinaus erlauben Mikrodaten die Überprüfung mikroökonomischer Hypothesen ohne restriktive Verteilungsannahmen. Besonders interessant sind für die empirische Wirtschafts- und Sozialforschung Paneldaten, die sich auf mehrere, meist in jährlichen Abständen erhobene Beobachtungen zu den einzelnen Beobachtungseinheiten (Personen, Haushalte, Betriebe, Unternehmen) beziehen und miteinander verknüpft werden können.
2. In wissenschaftlichen empirischen Untersuchungen mit Mikrodaten haben Personen- oder Haushaltsdaten im Vergleich zu Betriebs- und Unternehmensdaten ein deutliches Übergewicht. Dies ist auf das größere Angebot und die wesentlich einfacheren Zugangsmöglichkeiten im Vergleich zu Betriebs- oder Unternehmensdaten zurückzuführen. Mit deren einfacherer Verfügbarkeit für Forschungszwecke wird auch das Gewicht von Betriebs- und Unternehmensdaten zunehmen.

3. Für die empirische Wirtschaftsforschung in Deutschland ist die Verfügbarkeit von Mikrodaten von elementarer Bedeutung. Zum einen setzt eine empirisch fundierte Beratung zu den meisten wirtschaftspolitisch relevanten Fragestellungen in den Bereichen Erwerbstätigkeit und Einkommen qualitativ hochwertige Mikrodaten voraus. Zum anderen ist deren Verfügbarkeit auch eine wesentliche Voraussetzung dafür, die Wettbewerbsfähigkeit der empirischen Wirtschaftsforschung zu gewährleisten und den Wissenschaftlern die Möglichkeit zu bieten, mit deutschen Daten in referierten internationalen wissenschaftlichen Zeitschriften zu publizieren. Dies kann nur dann erwartet werden, wenn auch für Deutschland Mikrodatsätze in vergleichbarer Qualität wie für die im Bereich der empirischen Wirtschafts- und Sozialforschung dominierende USA zur Verfügung stehen.
4. Der Großteil der in referierten wissenschaftlichen Fachzeitschriften in den letzten zehn Jahren veröffentlichten empirischen Untersuchungen in den Bereichen Erwerbstätigkeit und Einkommen für Deutschland basiert auf Mikrodaten (Haushalts-, Personen- oder Firmendaten). Fast die Hälfte dieser Veröffentlichungen basiert auf dem Sozioökonomischen Panel (SOEP). Dies ist im Wesentlichen auf den wenig restriktiven Zugang und die aktive „Vermarktung“ dieser Datenbasis zurückzuführen. Mikrodaten der Amtlichen Statistik (einschließlich der Bundesanstalt für Arbeit) wurden bisher nur in geringem Umfang für wissenschaftliche Forschungszwecke verwendet. Der Hauptgrund dafür besteht in den sehr restriktiven Zugangsmöglichkeiten zu diesen Daten in der Vergangenheit.
5. In den letzten Jahren haben sich allerdings die Zugangsmöglichkeiten zu faktisch anonymisierten Personen- und Haushaltsdaten der Amtlichen Statistik (Mikrozensus, Einkommens- und Verbrauchsstichprobe, Zeitbudgeterhebung und den ersten drei Wellen der deutschen Stichprobe des Europäischen Haushaltspanels) auf dem Gebiet Erwerbstätigkeit und Einkommen deutlich verbessert. Das Institut für Arbeitsmarkt und Berufsforschung (IAB) der Bundesanstalt für Arbeit hat dazu mit der Bereitstellung von zwei Stichproben der sogenannte Beschäftigtenstatistik sowie der Leistungsstatistik – der sogenannten I<sub>A</sub>BS – einen weiteren wichtigen Beitrag geleistet. Die verbesserten Zugangsmöglichkeiten zu Mikrodaten wurden durch finanzielle Förderung der Deutschen Forschungsgemeinschaft und anderen Wissenschaftsorganisationen ermöglicht. Allerdings existiert für die Finanzierung der Weitergabe faktisch anonymisierter Daten der Einkommens- und Verbrauchsstichprobe 1998 durch das Statistische Bundesamt zur Zeit keine entsprechende Finanzierung, der relativ hohe Bezugspreis ist für die meisten Wissenschaftler vermutlich prohibitiv. Dies ist bedauerlich, da diese Erhebung die wichtigste Datenbasis für wissenschaftliche Untersuchungen im Bereich Einkommen, Sparen und Verbrauch privater Haushalte in Deutschland darstellt.
6. Seit kurzem besteht für Wissenschaftler die Möglichkeit, im Rahmen von Kooperationsprojekten mit den Statistischen Landesämtern und mit dem Statistischen Bundesamt vor Ort Mikrodaten der Amtlichen Statistik für wissenschaftliche Zwecke zu nutzen. Dies wird vom Statistischen Bundesamt als eine Möglichkeit angesehen, auch unternehmensbezogene Daten, die höhere Anforderungen an die Anonymisierung als personenbezogene Daten stellen, für wissenschaftliche Auswertungen zugänglich zu machen. Zudem prüft das Statistische Bundesamt die Möglichkeit der faktischen Anonymisierung von Betriebs- und Unternehmensdaten. Im Vergleich zu den USA, wo es Wissenschaftlern bereits seit Mitte der achtziger Jahre möglich ist, in Forschungsk Kooperationen mit dem US Bureau of

Census die Mikrodaten „vor Ort“ zu nutzen, hat die Öffnung der Amtlichen Statistik spät stattgefunden.

7. Allgemein sind die Zugangsmöglichkeiten zu wichtigen nicht-amtlichen Mikrodatenbasen in Deutschland allerdings weiterhin unbefriedigend. Es gibt keine laufend aktualisierte Übersicht über den vorhandenen (mit öffentlichen Mitteln geförderten) Datenbestand von Forschungseinrichtungen auf dem Gebiet von Erwerbstätigkeit und Einkommen. Wir vermuten aber, dass nur ein kleiner Teil der mit öffentlichen Mitteln erhobenen, relativ einfach anonymisierbaren Personen- und Haushaltsdaten an das Zentralarchiv für empirische Sozialforschung (ZA) Köln, das als Clearing-Stelle fungiert, weitergegeben wird.
8. Öffentlich finanzierte Institutionen, wie das Bundesinstitut für Berufsbildung (BIBB) und die Hochschul-Informationen-System (HIS) GmbH, führen laufend detaillierte Erhebungen im Bereich Qualifikation, Ausbildung und Erwerbstätigkeit durch, die sich relativ einfach anonymisieren ließen. Bisher sind diese Daten für externe Wissenschaftler nicht oder nur mit erheblichen zeitlichen Verzögerungen zugänglich, so dass das Analysepotenzial dieser Daten nicht oder nur unzureichend genutzt werden kann oder von den Wissenschaftlern genutzt wird. Da es sich bei diesen Datensätzen größtenteils um Personendaten handelt, wäre eine faktische Anonymisierung nach dem vom Statistischen Bundesamt bei der Weitergabe von Mikrodaten (beispielsweise dem Mikrozensus) praktizierten Verfahren problemlos möglich. Die verzögerte Weitergabe von faktisch anonymisierten Daten, wie dies vom BIBB praktiziert wird, hat unseres Erachtens keine sachliche Begründung.
9. Die Zugangsmöglichkeiten zu dem seit 1993 mit öffentlichen Mitteln erhobenen IAB-Betriebspanel für externe Wissenschaftler sind stark eingeschränkt und werden daher nur in geringem Ausmaß genutzt. Die bisher diskutierten und auch in Ansätzen bereits praktizierten Zugangsmöglichkeiten (Einrichtung einer Schalterstelle, Vor-Ort-Rechnen, Bereitstellung von Dissertationsstellen) sind für die meisten interessierten Wissenschaftler nicht attraktiv. Diese Daten wurden daher von externen Wissenschaftlern bisher kaum genutzt.
10. Alle verfügbaren Mikrodatensätze weisen mehr oder weniger einschneidende Einschränkungen auf, so dass für verschiedene Fragestellungen jeweils unterschiedliche Datensätze mit in der Regel eingeschränktem Informationsgehalt Verwendung finden müssen. So enthält beispielsweise der Mikrozensus als die wichtigste Erhebung der Amtlichen Statistik zur Bevölkerung und Erwerbstätigkeit in Deutschland zwar detaillierte Angaben darüber, aber keine Angaben zum Erwerbseinkommen. Dieses wird in der Einkommens- und Verbrauchsstichprobe relativ genau erfasst, diese Datenbasis enthält aber kaum Angaben zur Erwerbstätigkeit. Für die meisten wirtschaftswissenschaftlichen Fragestellungen im Bereich Erwerbstätigkeit und Einkommen sind aber detaillierte Angaben sowohl zur Erwerbstätigkeit als auch zum Einkommen erforderlich.
11. Von der Amtlichen Statistik werden verschiedene Datensätze mit detaillierten Informationen zu einzelnen Teilaspekten im Bereich Erwerbstätigkeit und Einkommen erhoben. Diese Datensätze erfüllen aber nur beschränkt die Anforderungen für wissenschaftliche Analysen. Ein wichtiger Grund dafür ist, dass in den einzelnen Datensätzen jeweils für mikroökonomische Analysen im Bereich Erwerbstätigkeit und Einkommen zentrale Variablen fehlen, die zwar in anderen Datensätzen (teilweise) erhoben werden, die einzelnen verfügbaren Mikrodatensätze aber unverbunden nebeneinander stehen. Daher ist die Verknüpfung der relevanten Information aus zwei oder mehreren Datensätzen nicht oder nur

mittels statistischer Verfahren, die mit großen Unsicherheiten behaftet sind, möglich. Im Haushaltsbereich betrifft dies vor allem den Mikrozensus und die Einkommens- und Verbrauchsstichprobe. Im Unternehmensbereich betrifft dies insbesondere das Nebeneinander der Gehalts- und Lohnstrukturerhebung (GLS) und der Amtlichen Industriestatistik.

12. Das SOEP enthält zwar in beschränktem Umfang Informationen zu beiden Themenkomplexen, Analysen zu kleineren Teilgruppen sind aufgrund des relativ geringen Stichprobenumfangs aber nicht oder nur sehr eingeschränkt möglich. Beispielsweise sind disaggregierte Auswertungen zu Beschäftigungsstruktur und Qualifikation nach Branchen aufgrund der geringen Fallzahlen nur sehr eingeschränkt möglich bzw. mit einer erheblichen statistischen Unschärfe behaftet. Detaillierte Analysen zur Struktur der Einkommen mittels des SOEP sind nur beschränkt möglich, da bestimmte Einkommenskomponenten (bspw. aus Kapitalvermögen oder Unternehmertätigkeit) nicht oder nur sehr unvollständig erfasst sind. Zudem ist der Anteil fehlender Angaben bei wichtigen Variablen relativ hoch.
13. Die wichtigsten verfügbaren Mikrodatsätze der Amtlichen Statistik (Mikrozensus, Einkommens- und Verbrauchsstichprobe, Gehalts- und Lohnstrukturerhebung) sind als reine Querschnitterhebungen angelegt. Damit entsprechen sie nur eingeschränkt den Anforderungen der wissenschaftlichen Forschung, die zunehmend auf Paneldaten basiert, da damit im Vergleich zu reinen Querschnittsdaten nicht nur die Heterogenität zwischen den Untersuchungseinheiten (Personen, Haushalten, Firmen) abgebildet werden kann, sondern auch dynamische Prozesse und Ungleichgewichtssituationen modelliert werden können. Im Bereich von Haushalts- und Personendaten liegt mit dem SOEP eine für viele Fragestellungen geeigneter Paneldatensatz vor. Die bisher wenig restriktiven Zugangsmöglichkeiten haben zu einer weiten Verbreitung (auch international) dieser wichtigen Datenbasis beigetragen und – wie bereits oben erwähnt – zu einer relativ großen Anzahl von Veröffentlichungen in referierten wissenschaftlichen Fachzeitschriften geführt. Mit dem IAB-Betriebspanel liegt mittlerweile zwar auch ein wichtiger Paneldatensatz mit Informationen zu Erwerbstätigkeit und Einkommen auf der Ebene einzelner Firmen vor. Diese Datenbasis konnte aufgrund der restriktiven Zugangsbeschränkungen von externen Wissenschaftlern bisher aber kaum genutzt werden.
14. Aus wissenschaftlicher Sicht ist das Analysepotenzial von administrativen Daten (Sozialhilfestatistik, steuerstatistische Einzeldaten) bei weitem noch nicht ausgeschöpft. Bislang gibt es nur wenige Pilotprojekte. Das Statistische Bundesamt hat in einer Reihe von Publikationen auf das Analysepotenzial dieser Daten hingewiesen.
15. Generell besteht ein Mangel an unterjährigen Mikrodaten, insbesondere bei personen- und haushaltsbezogenen Erhebungen. Dies betrifft vor allem auch den Mikrozensus als die wichtigste allgemeine Erhebung zur Erwerbstätigkeit in Deutschland. Im Statistischen Bundesamt bestehen zur Zeit Überlegungen der Ausweitung des Mikrozensus auf eine unterjährige Erhebung. Dies würde auch die Möglichkeit eröffnen, nach dem Rotationsprinzip Personen innerhalb eines kürzeren Zeitraums mehrmals zu befragen. Damit stünde beschränkt auch Panelinformation im Rahmen des Mikrozensus zur Verfügung.
16. Die Bundesanstalt für Arbeit verfügt über detaillierte prozess-produzierte Datenbestände (Beschäftigten-, Leistungsbezieher- und Förderstatistik), die bisher der Wissenschaft nicht oder nur ansatzweise zur Verfügung stehen. Einen wichtigen ersten Schritt dazu stellt die

oben erwähnte Bereitstellung von zwei 1%-Stichproben der Beschäftigten- und Leistungsstatistik (I<sub>A</sub>BS) dar. Allerdings wurden die zwei bisher erstellten Stichproben erst mit erheblicher zeitlicher Verzögerung verfügbar, so dass sie für Analysen zu aktuellen Fragestellungen von beschränktem Interesse sind. Die Bereitstellung von faktisch anonymisierten Stichproben aus der Förderstatistik und der Leistungsstatistik wären für die Wissenschaft ebenfalls von großem Interesse. Dies gilt auch für Stichproben aus der sogenannten Betriebsdatei der BA, die die personenbezogenen Mikrodaten der Beschäftigtenstatistik auf Betriebsebene aggregiert. Während die Weitergabe von personenbezogenen Daten der Beschäftigten-, Leistungsbezieher- und Förderstatistik in faktisch anonymisierter Form entsprechend dem vom Statistischen Bundesamt praktizierten Verfahren ohne datenschutzrechtliche Probleme möglich ist, ergeben sich bei betriebsbezogenen Daten die bereits oben erwähnten Probleme, die alternative Formen der Nutzung dieser Daten erfordern.

17. Vernetzungsmöglichkeiten zwischen Betriebs- und Personendaten wurden in anderen Ländern viel früher genutzt. In Teilbereichen, wie beispielsweise der betrieblichen Beschäftigungs- und Einkommensentwicklung, ist die Forschung in Deutschland nicht mehr wettbewerbsfähig. Forschung auf diesem Gebiet findet in Deutschland daher kaum statt. Die Versuche des IAB, personenbezogene Daten der Beschäftigtenstatistik mit dem IAB-Betriebspanel zu verknüpfen, sind daher zu begrüßen. Es muss allerdings gewährleistet sein, dass diese Daten auch externen Wissenschaftlern in geeigneter Form zur Verfügung stehen.
18. Für detaillierte Analysen zu bestimmten Fragestellungen im Bereich Erwerbstätigkeit und Einkommen stehen nur sehr eingeschränkt geeignete Informationen zur Verfügung, da die Amtliche Statistik die dafür notwendige Informationen nicht enthält oder die Stichproben der Datensätze mit entsprechender Information zu klein oder für wissenschaftliche Analysen nicht verfügbar sind. Dies betrifft beispielsweise Angaben zu den Einkommen von Hochqualifizierten nach Studienabschlüssen, die Erwerbstätigkeit in neuen Dienstleistungsberufen (z.B. im IT-Bereich), Angaben zur Erwerbstätigkeit und zum Einkommen auf der Ebene von Kreisen und Arbeitsamtsbezirken, Informationen zur „geringfügigen Beschäftigung“ und anderen Erwerbsformen, die nicht dem „Standard-Arbeitsverhältnis“ entsprechen und Mikrodaten zur Analyse der durch die Bundesanstalt für Arbeit geförderten Maßnahmen der „aktiven“ Arbeitsmarktpolitik.
19. Aufgrund des Mangels an informativen und frei verfügbaren Daten zu einzelnen Themenschwerpunkten in den Bereichen Erwerbstätigkeit und Einkommen werden von wissenschaftlichen Instituten und einzelnen Wissenschaftlern Daten mit meist sehr kleinen Stichproben mehrfach erhoben. So verdienstvoll und aufgrund der gegebenen Datenlage notwendig diese Befragungen auch sein mögen, sie können keinen Ersatz für eine breite nach einheitlichen Konzepten und kontinuierlich erhobenen Datenbasis sowohl für Personen und Haushalte als auch Betriebe und Unternehmen darstellen. Darüber wird durch die weiter zunehmende Befragungsintensität bei Haushalten und Unternehmen die Antwortbereitschaft massiv beeinflusst, worunter die Rücklaufquoten und die Datenqualität in erheblichem Ausmaß leiden.
20. Obwohl der Schwerpunkt dieser Expertise eindeutig auf Mikrodaten liegt, da diese für die wissenschaftliche Forschung im Bereich Erwerbstätigkeit und Einkommen von zentraler

Bedeutung sind, muss doch auch auf einige gravierende Mängel in der Verfügbarkeit aggregierter Daten der Amtlichen Statistik hingewiesen werden. Die fehlende Dienstleistungsstatistik ist die bedeutsamste Lücke im wirtschaftsstatistischen System Deutschlands. In einigen Bereichen der Wirtschaft, die für die zukünftige Entwicklung der deutschen Volkswirtschaft von zentraler Bedeutung sind, wie bspw. in der Informations- und Kommunikationsbranche, sind kaum Statistiken zu Erwerbstätigkeit und Einkommen verfügbar. Die „geringfügige“ Beschäftigung wird in der Amtlichen Statistik auf der Basis des Mikrozensus berechnet und damit nach allgemeiner Auffassung deutlich unterschätzt. Es fehlen auch Statistiken zu Arbeitsformen, die nicht dem „Normalarbeitsverhältnis“ entsprechen, aber von zunehmender Bedeutung sind, wie Leiharbeit, befristete Beschäftigungsverhältnisse und Selbständigkeit. Die in Tabellenform publizierten Daten der Amtlichen Statistik, bspw. aus dem Mikrozensus oder der Einkommens- und Verbrauchsstichprobe, erlauben keine tiefer gehenden Struktur- oder Verteilungsanalysen, da nur Kreuzklassifikationen mit wenigen Merkmalen nachgewiesen werden. Die Strukturhebung der Bundesanstalt für Arbeit ist zwar bezüglich der Struktur der Arbeitslosigkeit relativ detailliert, wird aber nur einmal pro Jahr durchgeführt. Sonderauswertungen sind zwar möglich, benötigen in der Regel jedoch viel Zeit und sind relativ teuer. Detailliertere Statistiken, insbesondere auf Basis des Mikrozensus, der Einkommens- und Verbrauchsstichprobe sowie Beschäftigtenstatistik, stehen im Internet bisher kaum zur Verfügung.

## Empfehlungen

Aus unserer Bestandsaufnahme der für wissenschaftliche Zwecke verfügbaren Daten im Bereich Erwerbstätigkeit und Einkommen ergeben sich aus unserer Sicht folgende Empfehlungen für eine Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik in diesem Bereich:

1. Da für Deutschland eine repräsentative Datenbasis der Amtlichen Statistik fehlt, die – vergleichbar zum Current Population Survey (CPS) in den USA – sowohl detaillierte Angaben zur Erwerbstätigkeit als auch zum Einkommen enthält, besteht hier akuter Handlungsbedarf. Als absolut vorrangig betrachten wir die Aufnahme zumindest des monatlichen *Bruttoerwerbseinkommens als metrisch gemessene Variable* (in DM bzw. EURO) in das Standardprogramm des Mikrozensus. Darüber hinaus sollten in beschränktem Umfang auch Angaben zu Sonderzahlungen, Transfers und dem Nettoerwerbseinkommen erhoben werden. Die dadurch entstehenden zusätzlichen Erhebungskosten könnten an anderer Stelle eingespart werden. Beispielsweise könnte die Beobachtungszahl und der Fragekatalog der Lohn und Gehaltsstrukturhebung reduziert werden.
2. Sehr begrüßenswert sind auch Überlegungen einer Arbeitsgruppe im Statistischen Bundesamt der unterjährigen (quartalsmäßigen) Erhebung des Mikrozensus, wobei - wie auch bei den Mikrozensuserhebungen anderer Länder üblich – nach dem Rotationsprinzip jedes Quartal ein Teil der Population ausgetauscht werden könnte. Durch die entstehende Panelstruktur wird der Mikrozensus für viele ökonomische Analysen interessant. Beispielswei-

se könnten dynamische Analysen zu Erwerbstätigkeit und Einkommen durchgeführt werden. Durch die Panelstruktur wird der Mikrozensus insbesondere auch für die empirische Arbeitsmarktforschung einen vergleichbaren Stellenwert wie der CPS in den USA erlangen.

3. Für eine detaillierte Erfassung der Einkommen sollte die Einkommens- und Verbrauchsstichprobe fortgeführt werden, allerdings auf jährlicher Basis. Obwohl wir den damit verbundenen Aufwand nicht verkennen, müsste dies durch neuere Entwicklungen in der Datenerhebungs- und -verarbeitungstechnologie organisatorisch möglich sein. Außerdem könnten aus wissenschaftlicher Sicht die laufenden Wirtschaftsrechnungen entfallen. Allerdings sollten auch in der Einkommens- und Verbrauchsstichprobe einige wichtige Angaben zur Erwerbstätigkeit, insbesondere auch zur Arbeitszeit erhoben werden.
4. Insgesamt ist eine stärkere Vernetzung der Amtlichen Statistik dringend geboten. Vielversprechend erscheinen hier Überlegungen einer beim Statistischen Bundesamt angesiedelten Arbeitsgruppe, den Mikrozensus als die Basis eines sogenannten „Access Panels“ für weitere Erhebungen zu machen. Diese Überlegungen sind aus wissenschaftlicher Sicht zu begrüßen, da sie in konsistenter Weise die Informationsbasis verbreitern. Sie sollten daher fortgeführt werden. Dabei muß aber gewährleistet sein, dass die Nutzungsmöglichkeiten für die Wissenschaft entsprechend berücksichtigt werden.
5. Die mit öffentlichen Mitteln erhobenen Daten im Bereich Erwerbstätigkeit und Einkommen gehören zur informationellen Infrastruktur eines modernen Staates und stellen somit zumindest hinsichtlich ihrer Verwertung für wissenschaftliche Zwecke ein öffentliches Gut dar. Die Nutzung dieser Daten sollte daher unter Wahrung der datenschutzrechtlichen Bestimmungen allen Wissenschaftlern und wissenschaftlichen Forschungseinrichtungen zu Grenzkosten für wissenschaftliche Zwecke ermöglicht werden. Um einen intensiven wissenschaftlichen Wettbewerb zu gewährleisten und Wettbewerbsbeschränkungen zu vermeiden, müssen diese Daten allen Wissenschaftlern zu den gleichen Bedingungen zugänglich sein. Nur so ist zu gewährleisten, dass die empirische Wirtschaftsforschung in Deutschland den Anschluss an das internationale Niveau nicht verliert.
6. Diese Einschätzung wird durch die positiven Erfahrungen mit dem Sozio-Ökonomischen Panel für Deutschland nachdrücklich belegt. Diese für wissenschaftliche Arbeiten im Bereich Erwerbstätigkeit und Einkommen zentrale Datenbasis, deren Finanzierung zur Zeit bei der Deutschen Forschungsgemeinschaft liegt, sollte auf alle Fälle fortgeführt und weiterhin in einer Form finanziert werden, die den freien Zugang zu dieser Datenbasis für alle Wissenschaftler gewährleistet.
7. Die Bereitstellung faktisch anonymisierter personenbezogener Mikrodaten der Amtlichen Statistik, wie dies vom Statistischen Bundesamt seit kurzem praktiziert wird, sollte unbedingt verlängert werden. Dies betrifft vor allem den Mikrozensus und die Einkommens- und Verbrauchsstichprobe, da seitens der Wissenschaft ein erhebliches Interesse an beiden Erhebungen besteht. Kurzfristig ist die Finanzierung der Weitergabe faktisch anonymisierter Daten der Einkommens- und Verbrauchsstichprobe 1998 an die Wissenschaft zu den gleichen Konditionen wie bei der EVS 1993 zu gewährleisten. Für viele Wissenschaftler an Universitäten ist die Anschaffung der EVS 1998 zu den zur Zeit angebotenen Preisen nicht finanzierbar. Langfristig sollte die Finanzierung dieses wichtigen Bestandteils der informationellen Infrastruktur Deutschlands durch das BMBF gewährleistet wer-

den. Sollten sich die Pläne der Umstellung des Mikrozensus und des Aufbaus eines Access Panels realisieren lassen, muss die Verfügbarkeit dieser Datenbasis für die Wissenschaft von Anfang an gewährleistet sein. Auch dazu bietet sich eine langfristige Grundfinanzierung durch das BMBF aus sachlichen Gründen an.

8. Die Bereitstellung faktisch anonymisierter personenbezogener Mikrodaten sollte auch auf andere mit öffentlichen Geldern finanzierte statistische Erhebungen ausgedehnt werden. Im Bereich der Amtlichen Statistik im engeren Sinn betrifft dies die Gehalts- und Lohnstrukturerhebung, die Amtliche Industriestatistik und die Lohn- und Einkommenssteuerstatistik. Dies betrifft aber auch die im Rahmen der Verwaltungsaufgaben generierten (prozess-produzierten) Daten der Bundesanstalt für Arbeit und die im Auftrag dieser Behörde oder über das ihr zugeordnete Institut für Arbeitsmarkt- und Berufsforschung durchgeführten Erhebungen von personenbezogenen Mikrodaten im Bereich Erwerbstätigkeit und Einkommen. Aus unserer Sicht ergeben sich hier datenschutzrechtlich keine anderen Gesichtspunkte als bei der faktischen Anonymisierung beispielsweise des Mikrozensus. Die Grundkosten der faktischen Anonymisierung sollten entweder ähnlich wie bspw. beim Mikrozensus durch das BMBF oder durch die Deutsche Forschungsgemeinschaft getragen werden.
9. Die Bereitstellung von faktisch anonymisierten personenbezogenen Daten der Bundesanstalt für Arbeit für Forschungszwecke soll fortgeführt werden. Dringend erforderlich ist die Bereitstellung faktisch anonymisierter Mikrodaten für die wissenschaftliche Evaluation arbeitsmarktpolitischer Maßnahmen durch die Bundesanstalt. Die dazu erforderlichen Daten sind bei der Bundesanstalt vorhanden und könnten in ähnlicher Form der Wissenschaft zur Verfügung gestellt werden wie zwei bereits verfügbare Stichproben aus der Beschäftigtenstatistik. Um eine möglichst aktuelle Datenbasis zu gewährleisten und die Personalkapazitäten der Bundesanstalt für Arbeit nicht zu belasten, sollte die Datenaufbereitung in Anlehnung an das beim Statistischen Bundesamt realisierte Konzept der Weitergabe faktisch anonymisierter Daten über eine Sonderfinanzierung erfolgen.
10. Die aus öffentlichen Mitteln finanzierten Erhebungen der Hochschul-Informationssystem (HIS) GmbH sollen der Wissenschaft schnellstmöglich in faktisch anonymisierter Form für Forschungszwecke zur Verfügung gestellt werden, da diese Erhebungen bisher von externen Wissenschaftlern kaum genutzt wurden und gerade in diesem Bereich ein akuter Mangel an verfügbaren Mikrodaten besteht. Methodisch sollten die Erfahrungen, die bei der faktischen Anonymisierung des Mikrozensus gewonnen wurden, berücksichtigt werden. Die Durchführung sollte von einem unabhängigen Institut mit einschlägiger Erfahrung übernommen werden. Die Finanzierung der faktischen Anonymisierung sollte das BMBF übernehmen.
11. Die BIBB/IAB Erhebung zu „Qualifikations- und Berufsverlauf“ wird von der Wissenschaft bisher zu wenig genutzt, da die Daten zum Zeitpunkt der Weitergabe veraltet sind und sich für Analysen zu aktuellen Fragestellungen nur beschränkt eignen. Die Nutzung dieser Daten könnte durch eine unverzügliche Weitergabe (d.h. nach Datenaufbereitung und Korrektur) vermutlich wesentlich intensiviert werden. Die mit der Durchführung dieser Erhebung beauftragten Institute sollten durch das BMBF verpflichtet werden, die Daten schnellstmöglich auch externen Wissenschaftlern in faktisch anonymisierter Form zur Verfügung zu stellen.

12. Allgemein sollten die mit öffentlichen Mitteln geförderten Erhebungen von personenbezogenen Mikrodaten nach Projektende beim Zentralarchiv für empirische Sozialforschung (ZA) oder einer anderen Clearingstelle für wissenschaftliche Analysen gelagert werden. Dies wird bei Erhebungen, die durch die Deutsche Forschungsgemeinschaft gefördert werden, bereits seit längerem praktiziert und hat sich gut bewährt. Damit würde auch der allgemeine Informationsstand über die Verfügbarkeit von Mikrodaten im Bereich Erwerbstätigkeit und Einkommen verbessert und Mehrfacherhebungen von Wissenschaftlern würden eingeschränkt werden.
13. Da die Methode der faktischen Anonymisierung bei Betriebs- und Unternehmensdaten in der Regel mit einem wesentlich größeren Informationsverlust verbunden ist als bei personenbezogenen Daten, sollten mehrere Möglichkeiten geprüft werden, wie mit öffentlichen Mitteln erhobene nicht-amtliche Betriebs- und Unternehmensdaten der Wissenschaft für Forschungszwecke zur Verfügung gestellt werden können. Da sich Erhebungen mit überwiegend qualitativen Fragen relativ einfach faktisch anonymisieren lassen und sich für eine Vielzahl von Analysen auch faktisch anonymisierte Datensätze mit starker Vergrößerung bestimmter Variablen (z.B. Umsätze, Beschäftigung) eignen, sollten entsprechende „Scientific Use Files“ zügig über das Zentralarchiv für empirische Sozialforschung oder einer anderen Clearingstelle für Forschungszwecke zur Verfügung gestellt werden. Bei den betriebs- und unternehmensbezogenen Daten der Amtlichen Statistik („Wirtschaftsdaten“) und anderen Erhebungen, die überwiegend quantitative Angaben enthalten, müssen auch andere Möglichkeiten geprüft werden, wie die Mikrodaten für Forschungszwecke zur Verfügung gestellt werden können.
14. Zu prüfen ist, ob durch organisatorische Änderungen die bestehenden Möglichkeiten für die Wissenschaft besser nutzbar gemacht werden können. Eine Möglichkeit dazu, die auch bereits in den USA und anderen Ländern praktiziert wird, sind dezentrale datenschutzrechtlich geschützte Rechenzentren. Eine weitere Möglichkeit stellt das „Fernrechnen“ über das Internet dar. Welche der Möglichkeiten zu einer bestmöglichen Nutzung der mit öffentlichen Mitteln erhobenen Daten auch für die Wissenschaft führt, sollte in einem bspw. durch das BMBF finanzierten Pilotprojekt näher geprüft werden.
15. Wichtige Grunddaten zur Erwerbstätigkeit und zum Einkommen sollten kostenlos im Internet zur Verfügung gestellt werden. Dazu zählen beispielsweise qualifikatorische Erwerbs- und Arbeitslosenquoten für verschiedene Altersgruppen und nach Geschlecht. Die Bereitstellung detaillierterer Kreuztabellierungen über das Internet durch das Statistische Bundesamt und die Bundesanstalt für Arbeit ist empfehlenswert und, wie ein Vergleich mit dem Datenangebot des U.S. Bureau of Labor Statistics zeigt, in Deutschland noch ausbaufähig.

# Inhaltsverzeichnis

<b>1</b>	<b>INFORMATIONSANBIETER UND ABGRENZUNG.....</b>	<b>1</b>
<b>2</b>	<b>DATENLAGE .....</b>	<b>3</b>
2.1	ÜBERBLICK ÜBER DAS VORHANDENE DATENANGEBOT .....	5
2.1.1	<i>Für die Wissenschaft frei verfügbare Mikrodaten.....</i>	<i>5</i>
2.1.2	<i>Für die Wissenschaft eingeschränkt verfügbare Mikrodaten .....</i>	<i>13</i>
2.1.3	<i>Für die Wissenschaft zur Zeit nicht verfügbare Mikrodaten .....</i>	<i>17</i>
2.1.4	<i>Aggregierte Daten .....</i>	<i>20</i>
2.2	DEFIZITE UND DATENBEDARF.....	23
2.2.1	<i>Defizite verfügbarer Mikrodaten.....</i>	<i>23</i>
2.2.2	<i>Defizite von aggregierten Daten.....</i>	<i>27</i>
2.3	VERNETZUNGSMÖGLICHKEITEN.....	29
2.4	ÜBERSCHNEIDUNGEN.....	31
<b>3</b>	<b>AUFTRAGGEBER FÜR DIE DATENPRODUKTION .....</b>	<b>32</b>
<b>4</b>	<b>FINANZIERUNG .....</b>	<b>33</b>
<b>5</b>	<b>RECHTSGRUNDLAGEN .....</b>	<b>35</b>
<b>6</b>	<b>DATENSCHUTZREGELUNGEN.....</b>	<b>35</b>
<b>7</b>	<b>INFORMATIONSVREBREITUNG.....</b>	<b>36</b>
7.1	INFORMATIONSVREBREITUNG ÜBER PRINTMEDIEN.....	37
7.2	INFORMATIONSVREBREITUNG ÜBER DAS INTERNET.....	37
<b>8</b>	<b>VERFÜGBARKEIT VON EINZELDATEN.....</b>	<b>38</b>
<b>9</b>	<b>WEITERGABE VON EINZELDATEN.....</b>	<b>40</b>
9.1	NATIONAL/INTERNATIONAL.....	40
9.2	PREISGESTALTUNG .....	42
9.3	WICHTIGE NUTZERGRUPPEN.....	42
	<b>LITERATURVERZEICHNIS .....</b>	<b>44</b>
	<b>ANHANG.....</b>	<b>51</b>



# 1 Informationsanbieter und Abgrenzung

Die wichtigsten Anbieter von Daten zur Erwerbstätigkeit und zum Einkommen für das gesamte Bundesgebiet sind das Statistische Bundesamt, die Bundesanstalt für Arbeit, das Institut für Arbeitsmarkt- und Berufsforschung in der Bundesanstalt für Arbeit (IAB) sowie Bundesinstitute wie z. B. das Bundesinstitut für Berufsbildung (BIBB), das der Rechtsaufsicht des BMBF unterstellt ist. Neben den amtlichen Datenproduzenten gibt es eine Vielzahl von Forschungsinstituten und –einrichtungen. Dazu zählen die im Rahmen der sogenannten „Blauen-Liste“-Institute geförderten Wirtschaftsforschungsinstitute und andere nicht-kommerzielle Wirtschaftsforschungsinstitute wie das Zentrum für Europäische Wirtschaftsforschung (ZEW) in Mannheim. Auch Stiftungen, wie z.B. die Hochschul-Informationssystem (HIS) GmbH, erheben laufend bundesweit Daten zur Arbeitsmarktsituation von Hochschulabsolventen. Daneben führen private Wirtschaftsforschungsinstitute und Unternehmensberatungen (bspw. Prognos und Kienbaum) mehr oder weniger regelmäßig Befragungen zu Erwerbstätigkeit und Einkommen von bestimmten Gruppen durch. Auch von wissenschaftlichen Instituten und einzelnen Wissenschaftlern werden manchmal die im Rahmen von wissenschaftlichen Projekten erhobenen Daten für Sekundärauswertungen nach Projektende zur Verfügung gestellt. Ein erheblicher Teil dieser Daten, einschließlich Fragebögen, wird vom Zentralarchiv für Empirische Sozialforschung (ZA) an der Universität zu Köln archiviert und an interessierte Wissenschaftler weitergegeben. Ein Großteil des Datenbestands bezieht sich auf sozialwissenschaftliche Erhebungen. Das Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim verfügt über anonymisierte Mikrodaten vom Statistischen Bundesamt, die für sozialwissenschaftliche Fragestellungen ausgewertet werden. Das Mannheimer Zentrum für Europäische Sozialforschung (MZES) verfügt mit ZEUS über eine Datenbank sozialwissenschaftlicher Daten für europäische Länder (Eurobarometer).

Der Bandbreite der Informationsanbieter entsprechend ist das Datenangebot in Deutschland sehr vielfältig.<sup>1</sup> Es ist daher im Rahmen dieses Beitrags unmöglich, das Datenangebot vollständig oder auch nur annähernd vollständig zu berücksichtigen. Eine Beschränkung der Darstellung auf die wichtigsten nach bestimmten Kriterien abgegrenzten Datenbestände ist daher unumgänglich. In dieser Expertise werden folgende Abgrenzungskriterien zugrunde gelegt:

- Die potenzielle Bedeutung der einzelnen Datenbasen für die wissenschaftliche Forschung im Bereich Erwerbstätigkeit und Einkommen. Dies ergibt sich aus der Problemstellung der vorliegenden Expertise und aus der Sicht empirisch arbeitender Wirtschaftsforscher als Nutzer von Daten im Bereich Erwerbstätigkeit und Einkommen. Daher werden hier nur Datensätze mit Schwerpunkten der Erhebung im Bereich Erwerbstätigkeit und Einkommen berücksichtigt. Datensätze, die zwar Angaben mit geringem Informationsgehalt zu Erwerbstätigkeit oder Einkommen enthalten, deren Schwerpunkte aber ganz überwiegend auf anderen Themenbereichen liegen, werden hier nicht berücksichtigt. Dies betrifft bspw. die im Auftrag des BMBF vom ZEW durchgeführte jährliche Unternehmensbefragung zum Innovationsverhalten in der deutschen Wirtschaft.

---

<sup>1</sup> Beispielsweise enthält der Datenbestandskatalog des ZA zu dem Stichwort Arbeitslosigkeit für den Erhebungszeitraum 1995 bis heute 41 Einträge. Nur ein Bruchteil der aufgeführten Datensätze entspricht den hier festgelegten Kriterien für die Aufnahme in den hier betrachteten Datenkatalog.

Andere Datensätze, auf die dieses Kriterium zutrifft, sind im Anhang zusammengefasst (vgl. Tabelle 5).

- Die Aufnahme in den Datenkatalog setzt voraus, dass die Daten für wissenschaftliche Analysen im Bereich Erwerbstätigkeit und Einkommen bereits zur Verfügung stehen oder aber es für derartige Analysen von Interesse wäre, wenn sie durch politische Entscheidungen der Wissenschaft prinzipiell zugänglich gemacht werden könnten. Privat erhobene Daten (bspw. von privaten Unternehmensberatungsfirmen) werden hier daher nicht betrachtet.
- Die Daten beziehen sich auf das gesamte Bundesgebiet. Im Anhang werden ergänzend auch wichtige regional beschränkte Datensätze erwähnt.
- Die Daten werden mehr oder weniger regelmäßig erhoben; Sonderbefragungen zu einem einzelnen Zeitpunkt werden nicht berücksichtigt. Nicht berücksichtigt werden auch mehrmalige Erhebungen, die bereits eingestellt wurden. In der Forschung verwendete Erhebungen, die nicht mehr fortgeführt werden, sind im Anhang erwähnt.
- Die Daten beziehen sich auf die Gesamtpopulation oder zumindest auf eine quantitativ bedeutsame Teilpopulation (z.B. Klein- und Mittelbetriebe, Hochschulabsolventen). Erhebungen für bestimmte Branchen oder sozio-ökonomische oder demografische Gruppen werden hier nicht betrachtet.

Insbesondere das erste Abgrenzungskriterium bedarf einer Präzisierung. Die potenzielle Bedeutung der Datensätze für die wissenschaftliche Forschung lässt sich objektiv nur schwer abschätzen. Was die Bedeutung der für die Wissenschaft bereits allgemein verfügbaren Datensätze anbelangt, ist ein Indikator dafür deren Verwendung bei wissenschaftlichen Publikationen in referierten Fachzeitschriften. Eine derartige Bewertung wurde beispielsweise kürzlich von Manser (1998) im Bereich der Arbeitsmarktforschung für die USA durchgeführt. Das Ergebnis dieser Untersuchung war, dass sich der Anteil der empirischen Arbeiten auf der Basis von Mikrodaten über die Zeit stark erhöht hat und dabei Personen- oder Haushaltsdaten im Vergleich zu Betriebs- und Unternehmensdaten deutlich dominieren. Der Hauptgrund für das starke Übergewicht von Studien auf der Basis von Personen- oder Haushaltsdaten wird von der Autorin in dem größeren Angebot und den wesentlich einfacheren Zugangsmöglichkeiten im Vergleich zu Betriebs- oder Unternehmensdaten gesehen.

Somit spiegelt die stärkere Verwendung von personen- oder haushaltsbezogenen Individualdaten nicht unbedingt das Forschungsinteresse der Wissenschaft, sondern vermutlich eher die unterschiedliche Datenverfügbarkeit relativ zu unternehmensbezogenen Individualdaten wider. Wie unsere eigene Auswertung der in referierten wissenschaftlichen Zeitschriften veröffentlichten Beiträge im Bereich Beschäftigung und Einkommen für Deutschland zeigt, repräsentiert die für die USA beschriebene Situation auch die relative Bedeutung von aggregierten relativ zu Individualdaten und von personenbezogenen relativ zu unternehmensbezogenen Daten in der deutschen Forschungslandschaft. Allerdings ist der Bestand an Individualdaten (insbesondere was personenbezogene Daten anbelangt), der für wissenschaftliche Zwecke nutzbar ist, in Deutschland wesentlich geringer als in den USA (gemessen am verfügbaren Datenbestand in einem Land).

Was die Datenverfügbarkeit anbelangt, ergibt sich ein Abgrenzungsproblem dadurch, dass die Auftraggeber regelmäßig erhobener Daten zwar häufig öffentliche Stellen (z.B. Ministerien oder die Bundesanstalt für Arbeit) sind, die Bereitstellung dieser Daten auch für wis-

senschaftliche Zwecke vom Auftraggeber in der Regel aber nicht als seine Aufgabe angesehen wird. Daher sind diese Daten für wissenschaftliche Zwecke meist nicht nutzbar. Dies trifft beispielsweise auf umfangreiche Datenbestände der Hochschul-Informationssystem GmbH zu, die durch den Bund und die Länder finanziert wird. Ein weiteres Beispiel dafür sind Daten der Bundesanstalt für Arbeit (Leistungsstatistik, Förderstatistik), die im Prozess der Arbeitsvermittlung und Arbeitsförderung anfallen. Da diese Expertise das Ziel verfolgt, allgemein auf Defizite bei der Bereitstellung von Daten für wissenschaftliche Zwecke im Bereich Erwerbstätigkeit und Einkommen hinzuweisen, werden auch Datenbestände, die ursprünglich nicht für diesen Zweck erhoben wurden, hier in die Betrachtung aufgenommen, sofern sie für die Wissenschaft von besonderem Interesse sind.

Im übrigen schließen wir uns der von Hauser, Wagner und Zimmermann (1998) in ihrem Memorandum „Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung“ formulierten Forderung an, nicht nur die Mikrodaten der amtlichen Statistik im engeren Sinne, sondern alle mit öffentlichen Mitteln erhobenen Daten der wissenschaftlichen Forschung zugänglich zu machen. Dies betrifft u.a. die im Rahmen der Auftragsforschung durch Ministerien, die Bundesanstalt für Arbeit und anderer Behörden erhobenen oder prozess-produzierten Mikrodaten.

## 2 Datenlage

Auf der Ebene von Personen, Haushalten, Betrieben und Unternehmen durch verschiedene Erhebungen gewonnene Daten (Mikrodaten) zur Erwerbstätigkeit und zum Einkommen bilden die Basis für die laufende Berichterstattung der amtlichen Statistik in diesem Bereich. Diese findet ihren Niederschlag in der Volkswirtschaftlichen Gesamtrechnung (VGR), den einzelnen Fachserien des Statistischen Bundesamtes und den Monatsberichten sowie jährlichen Strukturanalysen des Arbeitsmarktes der Bundesanstalt für Arbeit. Mikrodaten sind aber auch die Basis für wissenschaftliche Analysen im Bereich Erwerbstätigkeit und Einkommen. Generell bieten Mikrodaten aus der Sicht der empirischen Wirtschafts- und Sozialforschung eine Reihe von Vorteilen:

- Mikrodaten erlauben eine sachgerechte Abgrenzung der Untersuchungseinheiten. Die Aggregation der Untersuchungseinheiten kann nach den Anforderungen der Untersuchung vorgenommen werden.
- Der Informationsgehalt von Mikrodaten ist in der Regel ungleich höher als von aggregierten Daten. Besonders interessant sind für die empirische Wirtschafts- und Sozialforschung Paneldaten, die sich auf mehrere, meist in jährlichen Abständen erhobene Beobachtungen zu den einzelnen Beobachtungseinheiten (Personen, Haushalte, Betriebe, Unternehmen) beziehen. Wesentliche Vorteile gegenüber Querschnittsbetrachtungen sind die Möglichkeit der Abbildung dynamischer Prozesse und Ungleichgewichtssituationen sowie der Kontrolle unbeobachtbarer Heterogenität (zeitkonstante Firmen- und Individualeffekte).
- Im allgemeinen können mikroökonomische Hypothesen nur mittels Mikrodaten getestet werden, da durch die Aggregation von Beobachtungseinheiten strukturelle Parameter mit Verteilungsparametern vermischt werden. Bei Verwendung aggregierter Daten ist ein Rückschluss auf die ökonomisch interessanten Strukturparameter nur bei Kenntnis

der Verteilung der Merkmale in der Population möglich. Diese Verteilung kann nur auf der Basis von Mikrodaten geschätzt werden.

In den letzten 10 Jahren wurden bei ca. drei Viertel aller in referierten Fachzeitschriften veröffentlichten Untersuchungen im Bereich Erwerbstätigkeit und Einkommen für Deutschland Mikrodaten verwendet (siehe hierzu Tabelle 1). Gemessen an der Anzahl der Veröffentlichungen in referierten wirtschaftswissenschaftlichen Fachzeitschriften ist das Sozioökonomische Panel (SOEP) der mit großem Abstand wichtigste Datensatz.

Tabelle 1: Verbreitung von Mikro- und Makrodaten in referierten Fachzeitschriften auf dem Gebiet Erwerbstätigkeit und Einkommen (in vH)

	1990-1994	1995-1999	1990-1999
<b>Mikrodaten (Personen, Haushalts, Firmendaten)</b>			
SOEP (einschließlich LIS)	48	40	43
Nicht amtliche Firmenbefragungen <sup>a</sup>	16	24	20
IAB-Beschäftigtenstichprobe	0	6	4
EVS/Mikrozensus	0	5	3
Nicht amtliche Personenbefragungen (BIBB/IAB Erhebung, Arbeitsmarktmonitor)	1	4	3
<i>Aggregierte Daten (Gesamtwirtschaft, Branche, Region)</i>			
Statistisches Bundesamt/Bundesanstalt für Arbeit u.a.	34	20	26

*Anmerkungen:* Analysiert wurden 249 Fachzeitschriften auf dem Gebiet Erwerbstätigkeit und Einkommen (breite Abgrenzung) im Zeitraum 1990-1999 (Volltextrecherche)

<sup>a)</sup> Hierunter fallen das IAB-Betriebspanel, Hannoveraner Firmenpanel, IFO Unternehmenspanel, ZEW-Creditreform Datenbank, das Mannheimer Innovations Panel (MIP) und einmalige Erhebungen (beispielsweise der Firmendatensatz für die Metallindustrie, siehe Schnabel und Wagner 1992 und Kraft 1991).

Auf das SOEP entfallen mehr als 40 vH aller Veröffentlichung oder ca. 57 vH aller Veröffentlichungen auf der Basis von Mikrodaten. Allerdings steht diese Datenbasis der Wissenschaft schon seit wesentlich längerer Zeit zur Verfügung als die unten beschriebenen Mikrodatensätze der Amtlichen Statistik und anderer Anbieter. In der zweiten Hälfte der 90er Jahre hat sich der Anteil der Untersuchungen auf Basis von Mikrodaten von 66 auf 80 Prozent erhöht. Damit wird ein Trend nachvollzogen, der sich in den USA, die in der wirtschaftswissenschaftlichen, insbesondere empirischen Forschung eine dominierende Stellung einnehmen, schon seit geraumer Zeit abgezeichnet hat (vgl. z.B. Manser 1998). Die zunehmende Verbreitung von Mikrodatenanalysen in der zweiten Hälfte der 90er ist sowohl durch die Bereitstellung faktisch anonymisierter Personen- und Haushaltsdatensätze (IAB-Beschäftigtenstichproben, Mikrozensus und anderer nicht-amtlicher Personenbefragungen, siehe unten) als auch durch die zunehmende Verfügbarkeit von Firmenbefragungen (IAB-Betriebspanel, Hannoveraner Firmenpanel, ZEW-Firmendaten u.a., siehe Anhang) begünstigt worden. Allerdings

sind die Zugangsmöglichkeiten zu Mikrodaten in Deutschland, insbesondere was Firmendaten anbelangt, im internationalen Vergleich nach wie vor stark eingeschränkt.

Die Bereitstellung bisher nicht zugänglicher Personen- und Haushaltsdaten der statistischen Ämter und anderer Informationsanbieter sowie bessere Zugangsmöglichkeiten zu betriebs- und unternehmensbezogenen Daten („Wirtschaftsdaten“) würde der empirischen Wirtschaftsforschung eine Reihe von Analysen und Studien ermöglichen, für die ihnen bisher die Datengrundlagen fehlen. Zudem könnten Mikrodaten in Einzelfällen zügiger bereitgestellt werden. Damit sind die Voraussetzungen gegeben, dass auf dem Gebiet der empirischen Wirtschaftsforschung der wissenschaftliche Abstand zu den USA nicht weiter zunimmt.

## **2.1 Überblick über das vorhandene Datenangebot**

Im folgenden werden entsprechend der obigen Abgrenzung Datensätze beschrieben, die für wissenschaftliche Zwecke als Mikrodaten frei verfügbar sind oder der Wissenschaft eingeschränkt zur Verfügung stehen. Ergänzend werden auch Datensätze benannt, die der Wissenschaft bisher zwar noch nicht zur Verfügung stehen, die aber aus unserer Sicht schnellstmöglich auch externen Wissenschaftlern zugänglich gemacht werden sollten. Abschließend werden die allgemein verfügbaren Daten zu Erwerbstätigkeit und Einkommen der Amtlichen Statistik zusammengefasst, die auf aggregierten Daten der unten genannten Mikrodatensätze basieren. Diese sind für deskriptive Darstellungen hilfreich, da leicht verfügbar, für wissenschaftliche Zwecke aufgrund des hohen Aggregationsgrads aber nur sehr beschränkt brauchbar.

### **2.1.1 Für die Wissenschaft frei verfügbare Mikrodaten**

In diesem Abschnitt werden die für wissenschaftliche Zwecke frei verfügbaren Mikrodatensätze nach der obigen sachlichen Abgrenzung kurz beschrieben. „Frei verfügbar“ bedeutet hier, dass die Daten von einzelnen Wissenschaftlern oder unabhängigen wissenschaftlichen Einrichtungen gegen eine geringe Gebühr nach bestimmten datenschutzrechtlichen Regelungen (vgl. dazu Abschnitt 6) extern genutzt werden können.

#### **Sozio-ökonomisches Panel (SOEP)**

Das SOEP ist eine jährliche Wiederholungsbefragung von in Deutschland lebenden Personen außerhalb von Anstalten. Die Stichprobe umfasste 1998 etwa 6.600 Haushalte mit mehr als 12.700 Personen. Seit 1990 sind auch die neuen Bundesländer und Ostberlin in die Befragung einbezogen. Das SOEP enthält relativ detaillierte Informationen zur Erwerbstätigkeit und zum Einkommen. Eine wichtige Besonderheit des SOEP besteht darin, dass diese Information nicht nur zum Vormonat des Befragungszeitraums, sondern auch in Form sogenannter Monatskalendarien für die einzelnen Monate des vorhergehenden Kalenderjahres erhoben wird. Damit bietet das SOEP als einer der wenigen Datensätze die Möglichkeit, echte Verlaufsanalysen zur Erwerbstätigkeit durchzuführen. Für das Einkommen werden diese Kalendarien-

Angaben leider nicht mehr erhoben, wodurch die Analysemöglichkeiten stark eingeschränkt sind. Ein wichtiges Merkmal des SOEP besteht in der Möglichkeit der Verknüpfung von Personen- und Haushaltsinformation. Da das SOEP über einen sehr umfangreichen Fragenkatalog verfügt, wurde es für eine Vielzahl wirtschafts- und sozialwissenschaftlicher Fragestellungen als Datenbasis verwendet.<sup>2</sup>

Die Vorteile des SOEP sind vor allem die leichte Verfügbarkeit, der mittlerweile relativ lange Erhebungszeitraum (seit 1984 für die alten, seit 1990 für die neuen Bundesländer) sowie ein über die Zeit mehr oder weniger konstanter Kern von Fragen. Darüber hinaus werden in jeder Welle detaillierte Schwerpunktbefragungen zu aktuellen Themen durchgeführt, wodurch relativ aktuelle Daten zu wichtigen Forschungsfragen verfügbar sind. Zum Teil werden diese Themenschwerpunkte in unregelmäßigen Abständen wiederholt (beispielsweise zum Themenschwerpunkt Weiterbildung), so dass auch Vergleiche über die Zeit möglich sind.

Der größte Nachteil des SOEP besteht darin, daß Auswertungen zu kleineren Teilpopulationen aufgrund der Stichprobengröße nicht oder nur sehr eingeschränkt möglich sind. Beispielsweise sind Auswertungen zur Erwerbstätigkeit von Hochqualifizierten nach Studienabschlüssen nicht möglich. Auch sind Auswertungen zur Beschäftigungsstruktur und Qualifikation nach Branchen aufgrund der geringen Fallzahlen nur stark aggregiert (maximal ein Dutzend Branchen) möglich. Detaillierte Analysen zur Struktur der Einkommen sind mittels des SOEP auch nur sehr beschränkt möglich. Insbesondere sind bestimmte Einkommenskomponenten (bspw. aus Kapitalvermögen oder Unternehmertätigkeit) nicht oder nur sehr unvollständig erfasst.

Das SOEP zeichnet sich im Vergleich zu allen anderen für Deutschland verfügbaren Mikrodatsätzen durch eine sehr starke internationale Verbreitung aus. Es ist mittlerweile im Bereich Erwerbstätigkeit und Einkommen die bekannteste deutsche Datenbasis und wird von vielen Wissenschaftlern genutzt. Wie unsere Auswertung der in referierten Fachzeitschriften publizierten Beiträge gezeigt hat, stellt das SOEP die von der Wissenschaft mit Abstand am häufigsten verwendete Datenbasis dar.<sup>3</sup> Zum einen spielt hier sicher eine Rolle, dass das SOEP (für Westdeutschland) bereits seit 1984 existiert. Ein ganz wesentlicher Faktor für die starke Nutzung dieser für die empirische Forschung im Bereich Erwerbstätigkeit und Einkommen in Deutschland zentrale Datenbasis ist aber sicherlich auf die wenig restriktive Weitergabepolitik und das aktive Bemühen der Ausweitung des Nutzerkreises zurückzuführen. Dazu trägt auch das in einem Kooperationsprojekt mit einer Universität in den USA entwickelte SOEP-PSID „Equivalence-File“ bei (PSID steht für „Panel Study on Income Dynamics“). Dieses stellt aufbereitete Daten zu vergleichbaren Variablen für die beiden Länder, insbesondere im Bereich Erwerbstätigkeit und Einkommen, zur Verfügung. Dieser Datensatz ist auch deshalb von besonderem Interesse, da das SOEP nach dem PSID-Konzept entwickelt wurde. Ähnliche Projekte zur Erstellung von vergleichbaren Datensätzen sind zur Zeit für das Britische (BHPS) und für das Kanadische (SLID) Haushaltspanel in Bearbeitung, die Weitergabe dieser „Equivalence Files“ an die Wissenschaft ist für die nahe Zukunft angekündigt.

---

<sup>2</sup> Eine detaillierte Variablenbeschreibung und Angaben zu den Schwerpunktbefragungen in den einzelnen Wellen findet man unter <http://www.diw.de/soep/>.

<sup>3</sup> Vergleiche dazu auch das von der Projektgruppe Sozio-Ökonomisches Panel beim DIW zusammengestellte Verzeichnis der auf Basis des SOEP erstellten Publikationen unter <http://www.diw-berlin.de/soep/soeplit/>.

## **Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS)**

Mit dieser Befragung sollen Trends zur Erfassung des sozialen Wandels in Deutschland aufgezeigt werden. Jede Erhebung setzt sich aus einem Standardprogramm und verschiedenen Schwerpunkten zusammen. In den letzten Jahren waren die Schwerpunkte Soziale Ungleichheit und Wohlfahrtsstaat sowie Einstellungen gegenüber ethnischen Gruppen. Die repräsentative Erhebung wird seit 1980 in zweijährigem Abstand durchgeführt und umfasst ca. 3.600 Personen pro Welle. Auf dem Gebiet von Erwerbstätigkeit und Einkommen sind im Standardprogramm nur wenige Angaben verfügbar (persönliches Einkommen, Ausbildungsabschluß, Beruf). Die ALLBUS-Daten werden unmittelbar nach der Aufbereitung und Dokumentation allen interessierten Personen und Institutionen für Analysen zur Verfügung gestellt und können beim ZA bestellt werden.

## **Europäisches Haushaltspanel (ECHP)**

Das Europäische Haushaltspanel (European Community Household Panel, ECHP) ist eine europaweite Längsschnitterhebung der privaten Haushalte. Es wurde vom Statistischen Amt der Europäischen Gemeinschaft (Eurostat) initiiert. Es startete 1994 und wird jährlich erhoben. In Europa wurden etwa 60.000 zufällig ausgewählte Haushalte befragt, davon in Deutschland ca. 5.000 Haushalte. Die ersten drei Wellen des ECHP wurden vom Statistischen Bundesamt im Rahmen einer eigens dafür aufgebauten Panelstichprobe erhoben. Die Mikrodaten der ersten drei Wellen des ECHP können als faktisch anonymisierte Daten von unabhängigen wissenschaftlichen Einrichtungen in Deutschland vom Statistischen Bundesamt gegen eine geringe Gebühr bezogen werden. Ab der vierten Welle wird diese Erhebung vom Statistischen Bundesamt aber nicht mehr selbst durchgeführt, sondern im Rahmen des SOEP erhoben.

Zusätzlich zum individuellen Einkommen wird im ECHP das Haushaltseinkommen detailliert erhoben. Das ECHP eignet sich im Prinzip gut für internationale Vergleiche. Allerdings ist der Erwerb der Daten relativ teuer und die Möglichkeit des Erwerbs beschränkt sich auf die sogenannte „User Data Base“, bei dem die Information bezüglich einzelner Variablen im Vergleich zum „Production file“, das von den jeweiligen nationalen datenerhebenden Instituten an Eurostat geliefert wird, bereits „verdichtet“ und „harmonisiert“ wurde. So ist die Information zur Ausbildung in drei Kategorien zusammengefasst („less than second stage of secondary-level education“, „second-stage of secondary-level education“ und „third-level education“). Dadurch werden länderspezifische Besonderheiten verwischt, manchmal kommt es auch zu ziemlich unsinnigen Zusammenfassungen. Beispielsweise sind in der mittleren Kategorie Abitur und Lehre zusammengefasst, was noch angehen mag, in der höchsten Kategorie („third-level education“) findet sich aber neben der Fachhochschule und Universität auch die Fachschulausbildung, was wiederum wenig Sinn macht.

Eine weitere Einschränkung bezieht sich auf die Einkommensangaben. Während auf Haushaltsebene sowohl das Einkommen sehr detailliert erhoben und auch zwischen Netto- und Bruttoeinkommen unterschieden wird, liegen für die einzelnen Personen in einem Haushalt (mit Ausnahme Frankreichs) nur Angaben zum Nettoerwerbseinkommen vor. Jeder Per-

son in einem Haushalt wird zwar ein sogenannter „Brutto-Netto Faktor“ zugewiesen, dieser ist aber für jedes Haushaltsmitglied gleich und kann daher nicht zur Umrechnung des individuellen Nettoeinkommens in das entsprechende Bruttoeinkommen verwendet werden. Ein weiteres Problem der UDB des ECHP besteht darin, dass fehlende Einkommensangaben offenbar imputiert wurden, ohne dies kenntlich zu machen (bei den Einkommensangaben gibt es kaum fehlende Werte). Die Einkommensangaben sind daher mit Vorsicht zu interpretieren.

Da die Daten des ECHP für Deutschland erst seit kurzem allgemein für die Wissenschaft vorliegen und der Bezug dieser Daten für die anderen europäischen Länder über Eurostat mit hohen Kosten verbunden ist und darüber hinaus restriktiv gehandhabt wird, liegen unseres Wissens bisher keine wissenschaftlichen Veröffentlichungen dazu vor. Was die Verwendung der deutschen Daten des ECHP anbelangt, stehen diese in unmittelbarem Wettbewerb mit dem SOEP, das aufgrund des wesentlich längeren Erhebungszeitraums und der großen Verbreitung für Wissenschaftler erheblich attraktiver ist. Außer für Ländervergleichen besteht ein möglicher Vorteil des ECHP gegenüber dem SOEP unseres Erachtens zur Zeit höchstens in der detaillierteren Erfassung des Einkommens auf Haushaltsebene.

## **Mikrozensus**

Der Mikrozensus ist die amtliche Repräsentativstatistik über die Bevölkerung und den Arbeitsmarkt, an der jährlich 1% aller Haushalte in Deutschland beteiligt sind (laufende Haushaltsstichprobe).<sup>4</sup> Der Mikrozensus wird seit 1957 erhoben und dient der Bereitstellung statistischer Informationen über die wirtschaftliche und soziale Lage der Bevölkerung sowie über die Erwerbstätigkeit, den Arbeitsmarkt und die Ausbildung (Mehrzweckstichprobe).<sup>5</sup> Er schreibt die Ergebnisse der Volkszählung fort. Zudem dient er der Evaluierung anderer amtlicher Statistiken, wie z.B. der Einkommens- und Verbrauchsstichprobe (vgl. dazu unten). Die Stichprobenerhebung über Arbeitskräfte in der Europäischen Union (Arbeitskräftestichprobe der EU) ist in den Mikrozensus integriert. Die Ergebnisse des Mikrozensus bilden eine wichtige Informationsquelle für Regierung, Verwaltung, Wissenschaft und für die Bürger. Insgesamt nahmen im Jahr 1995 rund 370.000 Haushalte mit 820.000 Personen am Mikrozensus teil; darunter etwa 160.000 Personen in rund 70.000 Haushalten in den neuen Bundesländern und Berlin-Ost. Das Statistische Bundesamt führt die organisatorische und technische Vorbereitung des Mikrozensus durch. Die Durchführung der Befragung und die Aufbereitung obliegt den Statistischen Landesämtern (dezentrale Statistik).

Das Frageprogramm des Mikrozensus besteht aus einem festen Grund- und Ergänzungsprogramm mit jährlich wiederkehrenden Frageprogrammen, die überwiegend mit Auskunftspflicht belegt sind. Darüber hinaus gibt es in vierjährigem Rhythmus Zusatzprogramme, die teilweise von der Auskunftspflicht befreit sind. Das jährliche Grundprogramm des Mikrozensus umfaßt neben Angaben zur Erwerbstätigkeit und Einkommen u.a. Angaben zu persönlichen Merkmalen und zum Haushaltszusammenhang.

---

<sup>4</sup> Siehe hierzu [http://www.statistik-bund.de/micro/micro\\_c1.htm](http://www.statistik-bund.de/micro/micro_c1.htm)

<sup>5</sup> „Zweck des Mikrozensus ist es, statistische Angaben in tiefer fachlicher Gliederung über die Bevölkerungsstruktur, die wirtschaftliche und soziale Lage der Bevölkerung und der Familien, den Arbeitsmarkt, die berufliche Gliederung und Ausbildung der Erwerbsbevölkerung sowie die Wohnverhältnisse bereitzustellen.“ (Mikrozensusgesetz §1 (2)).

Die Vorteile des Mikrozensus bestehen vor allem in der großen Zahl an Beobachtungen und der systematischen und wiederkehrenden Erhebung eines umfangreichen Fragenkatalogs. Die große Anzahl an Beobachtungen erlaubt auch Analysen für kleine Gruppen. Beispielsweise können Beschäftigungsquoten von Hochschulabsolventen nach Studienrichtungen ausgewertet und auf die Gesamtpopulation hochgerechnet werden. Auch ist die Anzahl fehlender Werte aufgrund der Auskunftspflicht bei vielen Fragen relativ zu anderen Erhebungen gering.

Der wohl gravierendste Nachteil des Mikrozensus besteht in der Erfassung des Einkommens, das in der Höhe nicht nach den unterschiedlichen Einkommensarten klassifiziert und zudem nur in relativ groben Klassen erhoben wird. Einkommensanalysen auf der Basis des Mikrozensus sind daher nur sehr eingeschränkt – wenn überhaupt – möglich. Eine weitere Beschränkung des Mikrozensus besteht darin, dass die geringfügige Beschäftigung im Mikrozensus vermutlich deutlich untererfasst wird (siehe Schupp 1999).

Am 17. Januar 1996 ist das Mikrozensusgesetz modifiziert worden. Danach ergeben sich für eventuelle zukünftige Analysen folgende Änderungen. Die Informationen zur beruflichen Aus- und Weiterbildung werden wegen ihrer Bedeutung im jährlichen Rhythmus abgefragt. Zusätzlich sind Zweck der Maßnahme und dafür aufgewendete Stunden enthalten. Dafür wird das Jahr des höchsten beruflichen Ausbildungs- oder Hochschulabschlusses sowie die Hochschulfachrichtung letztmalig in 1996 und dann nur mehr im vierjährigen Rhythmus erhoben.

Bis vor kurzem konnten Individualdaten des Mikrozensus in faktisch anonymisierter Form vom Statistischen Bundesamt nur zu erheblichen Kosten bezogen werden.<sup>6</sup> Durch eine Finanzierung des BMBF konnte das Statistische Bundesamt faktisch anonymisierte Daten des Mikrozensus für wissenschaftliche Zwecke für die Jahre 1991, 1993, 1995, 1996 und 1997 zu sehr niedrigen Kosten (130 DM) zur Verfügung stellen. Damit liegt für die Wissenschaft erstmals die wichtigste amtliche Erhebung zur Erwerbstätigkeit in Form von faktisch anonymisierten Mikrodaten allgemein zugänglich vor. Der Mikrozensus enthält auch einen Block von Fragen im Rahmen der europäischen Arbeitserhebung von Eurostat. Bis 1996 waren diese Fragen für sonstige externe Nutzer nicht verfügbar. Seit 1996 ist die Arbeitserhebung Bestandteil der vom Statistischen Bundesamt für wissenschaftliche Zwecke weitergegebenen Daten.

Im Oktober 1998 hat die erste Mikrozensus Nutzerkonferenz stattgefunden. Ausgewählte Beiträge sind in einem Konferenzband zusammengefaßt worden (siehe Lüttinger 1999). Besonders geeignet ist der Mikrozensus für die Untersuchung kleiner Bevölkerungsgruppen, beispielsweise für selbständige Ausländer (siehe Baumann 1999), sowie Ingenieure und Naturwissenschaftler (siehe Pfeiffer und Falk 1999). Der Mikrozensus eignet sich aber auch für Analysen der Arbeitszeit (Wolf 1999).<sup>7</sup>

---

<sup>6</sup> Beispielsweise betragen die Kosten der vom ZEW beim Statistischen Bundesamt bestellten 70% Stichprobe des Mikrozensus mit ca. 70 Variablen für die Jahre 1991 und 1993 jeweils über DM 20.000.

<sup>7</sup> Auf Basis des Mikrozensus gibt es bislang nur wenige Artikel in referierten Zeitschriften. Eine der wenigen Ausnahmen ist Pfeiffer (1999) zu dem Zusammenhang von Selbständigkeit und Arbeitslosigkeit.

## **Einkommens- und Verbrauchsstichprobe**

Die Einkommens- und Verbrauchsstichprobe (EVS) wird seit 1962/1963 in Zusammenarbeit mit den statistischen Ämtern der Länder durchgeführt.<sup>8</sup> Im Bundesdurchschnitt werden 0,2% aller privaten Haushalte befragt. 1993 hat die erste Befragung in allen 16 Bundesländern stattgefunden. Die Teilnahme an der Einkommens- und Verbrauchsstichprobe ist freiwillig, 1993 waren rund 56 Tausend Haushalte an der Befragung beteiligt. Die aktuellste EVS liegt für das Jahr 1998 vor. Während früher jeder beteiligte Haushalt über das ganze Jahr Aufzeichnungen über seine Einnahmen und Ausgaben führte, nimmt seit 1998 jeder Haushalt nur mehr drei Monate teil (Rotationsverfahren).

Neben dem sogenannten Grundinterview, bei dem das monatliche Haushaltsnettoeinkommen in Klassen erfasst wird, werden in den Haushaltsbüchern das ganze Jahr über die Einnahmen und Ausgaben für alle Haushaltsmitglieder angeschrieben. Daraus lassen sich für die einzelnen Haushaltsmitglieder die Bruttoerwerbseinkommen und das Haushaltseinkommen berechnen, letzteres sowohl brutto als auch netto. Allerdings enthält die EVS keine Angaben zur Erwerbstätigkeit (Teilzeit, Vollzeit, geleistete oder bezahlte Arbeitsstunden).

Bis zur EVS 1993 konnten Individualdaten der EVS nur in besonderen Fällen für wissenschaftliche Zwecke verwendet werden. Durch eine Finanzierung des BMBF konnte das Statistische Bundesamt faktisch anonymisierte Daten der EVS erstmals für die Erhebung 1993 für wissenschaftliche Zwecke zur Verfügung stellen. Damit liegt für die Wissenschaft erstmals die wichtigste amtliche Erhebung zum Einkommen in Form von faktisch anonymisierten Mikrodaten allgemein zugänglich vor.

Veröffentlichungen zum Themenfeld Einkommensverteilung und Armutsmessung auf Basis der Mikrodaten der EVS liegen u.a. von Hauser (1999), Hauser und Becker (1999) und Becker (2000) vor.

## **Zeitbudgeterhebung 1991/92**

Die Zeitbudgeterhebung wurde vom Statistischen Bundesamt in Zusammenarbeit mit den statistischen Ämtern der Länder bei rund 7200 Haushalten durchgeführt (siehe Blanke et al. 1996). Die Befragung erfolgte im früheren Bundesgebiet und in den neuen Ländern via Quoten Auswahl zu vier Erhebungszeiträumen vom Herbst 1991 bis zum Sommer 1992. Sie ist in dieser Größenordnung (ca. 32000 Tagebücher) die bislang aktuellste Erhebung zur Zeitverwendung in Deutschland. Erfasst wurde die Zeitverwendung aller Haushaltsmitglieder ab dem 12. Lebensjahr. Diese sollten an zwei aufeinanderfolgenden Tagen ihren Tagesablauf im 5-Minuten-Rhythmus in einem Tagebuch mit eigenen Worten aufschreiben. Neben den im Vordergrund stehenden Hauptaktivitäten wurden auch Angaben zu gleichzeitigen Aktivitäten erfasst, zu den beteiligten bzw. anwesenden Personen, dem Ort der Hauptaktivität und für wen diese ausgeübt wird. Die in den Tagebüchern beschriebenen Haupt- und gleichzeitigen Aktivitäten wurden mit Hilfe einer aus über 200 Tätigkeiten bestehenden Aktivitätenliste verschlüsselt. Dazu gehören u.a. auch Erwerbstätigkeit/Arbeitssuche. Die Erhebung enthält allerdings keine Angaben zum Erwerbseinkommen, es wird nur das gesamte Nettoeinkommen für einzelne Haushaltsmitglieder verwendet. Bislang ist der Datensatz vor allem von Mitarbeitern

---

<sup>8</sup> Siehe hierzu [http://www.statistik-bund.de/micro/micro\\_c2.htm](http://www.statistik-bund.de/micro/micro_c2.htm)

des Statistischen Bundesamts verwendet worden (siehe u. a. Schwarz 1996 zur Verbreitung ehrenamtlicher Tätigkeiten).

Die faktisch anonymisierten Daten der Zeitbudgeterhebung wurden vom Statistischen Bundesamt ebenfalls in dem oben erwähnten vom BMBF finanzierten Projekt für wissenschaftliche Zwecke zur Verfügung gestellt. Nach den uns bekannten wissenschaftlichen Veröffentlichungen zu schließen, wurden diese Daten bisher kaum genutzt. Ein Grund dafür könnte natürlich sein, dass die Mikrodaten erst sehr kurzfristig verfügbar sind. Was das relativ geringe Interesse unter Wirtschaftswissenschaftlern anbelangt, könnte dies möglicherweise aber auch damit zusammenhängen, dass keine Lohnangaben verfügbar sind. Darüber hinaus wurden auch keine retrospektiven Angaben zur individuellen Erwerbsbiographie erhoben, was ebenfalls das Analysepotenzial dieser Datenbasis einschränkt. Es ist geplant, diese Erhebung 2001/2002 erneut durchzuführen.

### **IAB-Beschäftigtenstichprobe (I<sub>A</sub>BS)**

Die IAB-Beschäftigtenstichprobe (I<sub>A</sub>BS) ist eine einprozentige (anonymisierte) Zufallsstichprobe aller sozialversicherungspflichtig beschäftigten Arbeiter und Angestellten in der Bundesrepublik Deutschland. Datenbasis sind die von den Betrieben an die Sozialversicherungsträger gemeldeten Beschäftigungsinformationen. Nicht enthalten sind Beamte, Selbständige, mithelfende Familienangehörige. Untererfasst sind geringfügig Beschäftigte. Untersuchungsgebiet sind die alten Bundesländer einschließlich West-Berlin und seit 1992 Ostdeutschland einschließlich Ost-Berlin. Die I<sub>A</sub>BS enthält jeweils ca. 200000 sozialversicherungspflichtig Beschäftigte im Querschnitt und über den gesamten Zeitraum ca. 430000 Personen.

Eine Besonderheit der I<sub>A</sub>BS sind tagesgenaue Verlaufsangaben über das Beschäftigungsverhältnis. Zudem ist die Validität der von den Betrieben gemeldeten Bruttoeinkommen bis einschliesslich der Beitragsbemessungsgrenze sehr hoch. Durch Verknüpfung mit den anonymisierten Daten der Leistungsstatistik der Bundesanstalt für Arbeit können den einzelnen Personen auch Phasen registrierter Arbeitslosigkeit zugeordnet werden, so dass partiell individuelle Erwerbskarrieren konstruiert werden können (Phasen der Nichterwerbstätigkeit können nur residual erfasst werden). Im Vergleich zu (retrospektiv erhobenen) Befragungsdaten können dadurch Erwerbsverläufe relativ fehlerfrei erfasst werden. Eine weitere Besonderheit der I<sub>A</sub>BS besteht darin, dass die einzelnen Beschäftigten über eine anonyme Betriebsnummer dem jeweiligen Betrieb zugeordnet werden können. Damit hat diese Datenbasis das Potenzial für eine Verknüpfung von Betriebs- und Personendaten (siehe dazu unten). Darüber hinaus bietet die Beschäftigtenstatistik außerordentlich hohe Fallzahlen und ist auch für Analysen, die sich auf bestimmte Teilgruppen (beispielsweise in einem bestimmten Wirtschaftszweig oder einer Berufsgruppe beschäftigte Jugendliche) beziehen, geeignet.

Ein wesentlicher Nachteil dieser Datenbasis besteht darin, dass die Einkommensangaben bei der oberen Sozialversicherungsgrenze zensiert sind. Dies betrifft vor allem Beschäftigte mit Hochschulabschluss, für die bereits der Median zensiert ist. Die Analysemöglichkeiten dieser Datenbasis sind dadurch beschränkt, dass zwar zwischen Vollzeit und Teilzeit (in zwei Kategorien) unterschieden wird, aber keine Information zu den geleisteten Arbeitsstunden verfügbar ist. Ein weiterer Nachteil besteht darin, dass die Anzahl der vorhandenen Merkmale, insbesondere auch bezüglich des Haushaltszusammenhangs, sehr beschränkt sind.

Faktisch anonymisierte Daten der IAB-Beschäftigtenstichprobe für den Zeitraum 1975 – 1990 wurden in einem gemeinsam vom Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit (IAB), dem Wissenschaftszentrum Berlin für Sozialforschung (WZB) und dem Zentrum für Umfragen und Analysen (ZUMA) in Mannheim durchgeführten Projekt der Wissenschaft über das Zentralarchiv für empirische Sozialforschung (ZA) in Köln zur Verfügung gestellt. Diese Datenbasis wurde in zwei Varianten angeboten: Zum einen mit detaillierter Brancheninformation (Zweisteller) und im wesentlichen ohne Regionalinformation, zum anderen mit Regionalinformation (Bundesländer) aber ohne Brancheninformation. In einem von der Deutschen Forschungsgemeinschaft (DFG) finanzierten Nachfolgeprojekt und kürzlich abgeschlossenen Projekt wurden vom IAB vergleichbare Daten für den Zeitraum 1975 – 1995 aufbereitet und wiederum über das ZA der Wissenschaft zur Verfügung gestellt.

Die I<sub>A</sub>BS wird mittlerweile von mehr als 50 Wissenschaftlern genutzt. Diese Datenbasis eignet sich aufgrund der großen Beobachtungszahl insbesondere für Untersuchungen zu Erwerbstätigkeit und Erwerbseinkommen für bestimmte Gruppen oder disaggregiert nach Branchen oder Berufen. Auch sind mit dieser Datenbasis Panelanalysen möglich. Mittlerweile sind eine Reihe von Arbeiten auf Basis der I<sub>A</sub>BS erschienen. Beispiele dafür sind: Lohnstruktur (Möller 1999, Steiner und Wagner 1998), Arbeitslosigkeit (Mavromaras und Rudolph 1998), Arbeitskräftemobilität (Bender et al. 1999), Löhne und Arbeitslosigkeit (Baltagi und Blien 1998) sowie Außenhandel und Löhne (Fitzenberger 1999). Weitere Analysemöglichkeiten mit der I<sub>A</sub>BS 1975-1995 werden in Bender et al. (2000) beschrieben.

### **BIBB/IAB Erhebung „Qualifikation und Berufsverlauf“**

Diese Erhebung ist eine gemeinsame Erhebung vom Bundesinstitut für Berufsbildung (BIBB) und dem Institut für Arbeitsmarkt und Berufsforschung (IAB). Bisher wurden vier Erhebungen durchgeführt. Diese betreffen die Jahre 1978/79, 1985/86, 1991/92 und 1998/99. Die BIBB/IAB Erhebung enthält eine Vielzahl von Informationen zu Qualifikation und Erwerbstätigkeit (Alex 2000, Jansen 1993). Aufgrund der Vielzahl der offenen Fragen ist die Aufbereitung der BIBB/IAB Erhebung sehr zeitaufwendig. Die Aufbereitung dieses Datensatzes gilt als vorbildlich. Insbesondere werden Erwerbstätige mit einer abgeschlossenen Lehre nach der Verwertbarkeit ihrer Ausbildung gefragt. Der Vorteil dieser Datenbasis besteht in der großen Anzahl an Beobachtungen und der detaillierten Information zur Qualifikationsstruktur und deren Determinanten. Dadurch sind detaillierte Analysen auch zu relativ kleinen Gruppen von Erwerbstätigen (z.B. zu älteren Arbeitnehmern in bestimmten Wirtschaftssektoren) möglich. Eine Besonderheit sind die Informationen zu benötigten Arbeitsmitteln (z. B. Werkzeuge und Computer), physische und psychosoziale Anforderungen und Belastungen sowie organisatorischen und betriebsstrukturellen Rahmenbedingungen. Thematischer Schwerpunkt der jüngsten Erhebung ist der strukturelle Wandel der Arbeitswelt, seine Auswirkungen auf die Arbeitsbedingungen und auf das individuelle Mobilitätsverhalten. Ein weiterer Vorteil dieser Datenbasis besteht darin, dass der Großteil des Fragenkatalogs in den einzelnen Erhebungen konstant bleibt, wodurch auch zeitliche Vergleiche der Qualifikationsstruktur und deren Determinanten möglich sind. Die Ergebnisse werden vor allem in hausinternen Zeitschriften vorgestellt. Die Verbreitung dieses Datensatzes in der Wissenschaft ist vermutlich dadurch stark eingeschränkt, dass die Daten erst mit erheblicher zeitlicher Verzögerung verfügbar ge-

macht werden. Nach ca. 5 Jahren wird die BIBB/IAB Erhebung an das ZA weitergegeben. Zur Zeit können die ersten drei Erhebungen über das ZA bezogen werden.

In der empirischen Wirtschafts- und Sozialforschung ist der Datensatz für verschiedene Fragestellungen eingesetzt worden: Lohnstruktur und Computer (DiNardo und Pischke 1997), Determinanten der betrieblichen Ausbildung (Acemoglu und Pischke 1998), Verwertbarkeit der betrieblichen Ausbildung (Harhoff und Kane 1997, Blechinger und Pfeiffer 1998) und Übergang von Lehre in Beruf (Franz und Zimmermann 1999).

### **Lebensverlaufsstudien des Max-Planck-Instituts für Bildungsforschung**

Seit über zehn Jahren wird am Max-Planck-Institut für Bildungsforschung in Berlin im Forschungsbereich "Bildung, Arbeit und gesellschaftliche Entwicklung" unter der Leitung von Prof. Karl-Ulrich Mayer das Langzeitprojekt "Lebensverläufe und sozialer Wandel" durchgeführt (siehe Brückner und Mayer 1995).<sup>9</sup> Bislang wurden im Rahmen dieses Projekts knapp 8.000 Lebensverläufe von deutschen Staatsangehörigen unterschiedlicher Geburtskohorten in West- und Ostdeutschland mit einem weitgehend standardisierten Erhebungsinstrument erhoben. Die westdeutschen Teilstudien des Projekts "Lebensverläufe und gesellschaftlicher Wandel" sowie die erste ostdeutsche Teilstudie können über das ZA bezogen werden. Der Datensatz enthält von über 5.000 Männern und Frauen der Geburtsjahrgänge 1919-1921, 1929-1931, 1939-1941, 1949-1951, 1954-1956 und 1959-1961 detaillierte Lebensverlaufsinformationen. Zu Erwerbstätigkeit und Einkommen sind detaillierte Informationen verfügbar zu: Beginn und Ende aller Ausbildungen einschließlich der erreichten Qualifikationen, Beginn und Ende aller Erwerbstätigkeiten, Erwerbsunterbrechungen, berufliche Tätigkeiten, berufliche Stellungen, Arbeitszeit, Verdienst, verschiedene Einkommensquellen, persönliches Einkommen (im Querschnitt). Diese Längsschnittsdaten erlauben die Untersuchung zahlreicher Fragestellungen u.a. auch zu Erwerbstätigkeit und Bildungsverhalten sowie Einkommensanalysen.

#### **2.1.2 Für die Wissenschaft eingeschränkt verfügbare Mikrodaten**

Seit kurzem steht der Wissenschaft die Möglichkeit offen, im Rahmen von Kooperationsprojekten mit den Statistischen Landesämtern und mit dem Statistischen Bundesamt vor Ort mit den Mikrodaten der Amtlichen Statistik zu rechnen. Dies wird vom Statistischen Bundesamt als eine Möglichkeit angesehen, auch unternehmensbezogene Daten, die höhere Anforderungen an die Anonymisierung als personenbezogene Daten stellen (siehe dazu Abschnitt 6), für wissenschaftliche Auswertungen zugänglich zu machen. Zudem prüft das Statistische Bundesamt die Möglichkeit der faktischen Anonymisierung von Betriebs- und Unternehmensdaten. Dieses Projekt soll sowohl die amtliche Industriestatistik als auch die Kostenstrukturerhebung für Unternehmen miteinbeziehen. Im folgenden werden die Datensätze kurz beschrieben und auf mögliche Anwendungspotenziale im Bereich Erwerbstätigkeit und Einkommen hingewiesen. Aufgenommen werden hier Datensätze, die bereits von Wissenschaftlern ge-

---

<sup>9</sup> Siehe auch <http://www.mpib-berlin.mpg.de/BAG/Studien/ZA-Info38/Wagner.htm>.

nutzt werden konnten oder für die bereits Vorschläge für konkrete Anträge auf Kooperationsprojekte mit dem Statistischen Bundesamt vorliegen.

### **Gehalts- und Lohnstrukturerhebung (GLS)**

Die Gehalts- und Lohnstrukturerhebung (GLS) steht im Mittelpunkt des lohnstatistischen Berichtssystems und wird alle fünf Jahre auf der Personen- und Firmenebene durchgeführt (zuletzt 1995).<sup>10</sup> Die Stichprobenerhebung umfasst ca. 900 Tsd. Beschäftigte im Produzierenden Gewerbe, Handel und bei Banken und Versicherungen. Die GLS enthält auf der Personenebene Angaben über Verdienste (einschließlich aller tariflichen oder außertariflichen Leistungs-, Sozial- und sonstigen Zulagen oder verschiedene Sonderzahlungen, z. B. Gewinnbeteiligung) nach verschiedenen Merkmalen (Geschlecht, Ausbildungsstand, Familienstand, Lebensalter, ausgeübte Tätigkeit, Dauer der Unternehmenszugehörigkeit), über die Arbeitszeit (effektiv geleistete Stunden unabhängig von der Bezahlung, un- und bezahlte Mehrarbeitsstunden etc.) sowie Angaben auf der Firmenebene über die Beschäftigtenzahl der Unternehmen.

Die GLS ist eine potenziell wichtige Datenbasis für Analysen zur Erwerbstätigkeit und zum Erwerbseinkommen. Die GLS ist insbesondere auch für den Hochlohnbereich interessant, da beispielsweise in der I<sub>A</sub>BS, wie oben erwähnt, Entgeltangaben nur für sozialversicherungspflichtig Beschäftigte bis zur Bemessungsgrenze der Rentenversicherung ausgewiesen sind. Zudem sind in den bisher verfügbaren Mikrodatsätzen lediglich unbefriedigende Angaben über Löhne und Arbeitszeiten vorhanden (Mikrozensus: Nettolöhne, SOEP: geringe Fallzahl, viele Unplausibilitäten, I<sub>A</sub>BS: keine exakte Arbeitszeit). Genauigkeit der Angaben und Stichprobenumfang der GLS lassen wesentlich genauere und spezifischere Untersuchungen zu, beispielsweise über die Verteilung der Stundenlöhne nach Qualifikation und Branchen. Weitere Anwendungsbeispiele sind Wirkungen einer Verkürzung der Arbeitszeit und qualifikatorische Arbeitsnachfrage, insbesondere die Substitutionsbeziehungen verschiedener Gruppen von Arbeit. Bisher liegt unseres Wissens auf der Basis der GLS nur eine wissenschaftliche Untersuchung vor. Stephan (2000) untersucht die Determinanten der Lohnstruktur anhand der Niedersächsischen Gehalts- und Lohnstrukturerhebungen 1990 und 1995.

Nachteilig für wissenschaftliche Analysen ist, dass die GLS nicht zentral beim Statistischen Bundesamt, sondern dezentral vorliegt und empirische Untersuchungen sich auf das jeweilige Bundesland beziehen. Im Prinzip könnte die GLS in faktisch anonymisierter Form für wissenschaftliche Zwecke zur Verfügung gestellt werden. Die bei der faktischen Anonymisierung auftretenden Probleme dürften ähnlich wie beispielsweise beim Mikrozensus sein.

---

<sup>10</sup> Siehe Fachserie 16, Reihe 2.1-2.3.

## **Amtliche Industriestatistik für Betriebe und Unternehmen**

Für das Produzierende Gewerbe stehen eine Vielzahl von Betriebs- und Unternehmensdaten zur Verfügung. Der Monatsbericht im Bergbau und Verarbeitenden Gewerbe für Betriebe und Unternehmen umfasst Auftragseingänge, tätige Personen, Arbeitsstunden, Lohn- und Gehaltssummen, den Umsatz insgesamt und den Auslandsumsatz.<sup>11</sup> Teilweise sind Betriebsdaten zu einem Panel verknüpft worden. Vorteile gegenüber anderen Firmendatensätzen sind geringe Antwortausfälle, hohe Fallzahlen und gute Qualität der quantitativen Angaben. Nachteilig ist der aus der Sicht der Erwerbstätigkeit kleine Fragenkatalog.

Anfang der neunziger Jahre wurde in einem Gemeinschaftsprojekt zwischen dem Niedersächsischen Landesamt (NLS) und Wissenschaftlern der Universitäten Hannover und Lüneburg (Prof. Gerlach und Prof. Wagner) erstmals für Deutschland mit der Nutzung amtlicher Betriebsdaten durch die Wissenschaft begonnen (siehe Wagner 1994, 1999). Thema des Forschungsschwerpunkts sind Produktionsentwicklung, Beschäftigungswachstum und Exporterfolge niedersächsischer Industriebetriebe, sowie Arbeitsplatzdynamik und die mikroökonomische Arbeitsnachfrage (siehe hierzu Gerlach und Wagner 1992, 1993, Breitung 1994). Mittlerweile gibt es eine Vielzahl von Kooperationsprojekten zwischen Universitäten und den Statistischen Landesämtern.<sup>12</sup> Im Rahmen dieser Kooperationsprojekte bekommen Mitarbeiter von Universitäten einen "Mitarbeiter-Status" im Statistischen Landesamt und können dort direkt mit den Firmendaten arbeiten. Inhalte der Forschungsprojekte sind Beschäftigungseffekte von Firmenschließungen und Firmengründungen, Determinanten der Produktivitätsentwicklung und Arbeitsnachfrage (siehe Statistisches Bundesamt 1999b).

### **IAB-Betriebspanel**

Seit 1993 wird vom Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit (IAB) eine jährliche repräsentative Arbeitgeberbefragung von Betrieben durchgeführt (siehe hierzu Bellmann 1997).<sup>13</sup> Seit 1996 werden auch Betriebe in den neuen Bundesländern befragt. Grundgesamtheit des IAB-Betriebspanels sind alle Betriebe einschliesslich Dienstleistungen mit mindestens einem sozialversicherungspflichtig Beschäftigten. Die Ziehung der Zufallsstichprobe erfolgte auf Grundlage einer Betriebsdatei, die aus der Beschäftigtenstatistik der Bundesanstalt für Arbeit generiert wurde. Die Stichproben enthalten pro Befragungswelle jeweils über 4.000 Fälle für Ost- und Westdeutschland. Die siebte Welle West bzw. die vierte Welle Ost aus 1999 enthält insgesamt 9.762 Betriebe, wobei sich 5.335 Fälle auf die neuen Bundesländer und Ost-Berlin und 4.427 Fälle auf das frühere Bundesgebiet beziehen. Die Rücklaufquote der mündlichen Befragung beträgt außer für die erste Welle Ost immer über 70%. Für das Jahr 2000 wird das IAB-Betriebspanel mit Beteiligung einiger Bundesländer auf den doppelten Umfang aufgestockt. Das IAB-Betriebspanel wird zum Teil aus Mitteln des Europäischen Sozialfonds gefördert.

---

<sup>11</sup> Siehe Fachserie 4, Reihe 4.1.1.

<sup>12</sup> Siehe die website FiDASt Firmen-Daten aus der Amtlichen Statistik Nutzer-Netzwerk <http://www.uni-lueneburg.de/fb2/vwl/wifo/FiDASt/s-anhalt.htm>.

<sup>13</sup> Das IAB stellt auf Anfrage ein Codebook zum IAB-Betriebspanel zur Verfügung.

Das IAB-Betriebspanel ist potenziell eine wichtige Datenbasis auf Betriebsebene im Bereich der Erwerbstätigkeit. Der Fragenkatalog umfaßt nahezu alle Bereiche der Erwerbstätigkeit und eingeschränkt auch Fragen zum Einkommen. Bislang liegen eine Vielzahl von Analysen mit dem IAB-Betriebspanels vor. Diese betreffen beispielsweise Analysen zum Zusammenhang zwischen Arbeitsnachfrage und Innovation (Bellmann and Kohaut 1999), Qualifikatorische Arbeitsnachfrage (Bellmann et al. 1999), Arbeitszeit, Ausbildungsbeteiligung der Betriebe (Bellmann und Lahner 1998), Einfluß von Betriebsräten auf befristete und geringfügige Beschäftigung (Düll und Ellguth 1999). Im Bereich Einkommen werden betriebliche Determinanten der Lohnhöhe (Bellmann und Blien 1997), Angaben zur Tarifbindung und Verbreitung der Verbreitung von Flächentarifverträgen untersucht (Kohaut und Schnabel 1998). Nach Angaben des IAB ist es aus datenschutzrechtlichen Gründen bisher nicht möglich, die Mikrodaten des IAB-Betriebspanel externen Wissenschaftlern zur Verfügung zu stellen. Zur Zeit werden alternative Möglichkeiten der externen Nutzung dieser Daten erprobt (siehe dazu Abschnitt 6). Nach unserem Kenntnisstand sind bisher keine publizierten Analysen von externen Nutzern vorhanden.

### **Computerstützte Arbeitsvermittlung (CoArb)**

In der CoArb sind sämtliche vermittlungsrelevante Informationen zu arbeitslosen und arbeitssuchenden Personen gespeichert. Außerdem enthält die Datenbank alle den Arbeitsämtern gemeldeten offenen Stellen. Die CoArb unterliegt einer ständigen Aktualisierung. Bei jedem Beratungsgespräch fügt der Arbeitsvermittler der Datenbank neue Informationen über den Bewerber hinzu. Die offenen Stellen werden ebenfalls ständig aktualisiert. Die CoArb ist dezentral organisiert und umfasst alle Arbeitsamtsbezirke. Dies bedeutet, daß die in den einzelnen Arbeitsamtsbezirken vorgehaltenen Daten nicht über die Grenzen der Bezirke hinweg ausgetauscht werden können.<sup>14</sup> Allerdings besteht eine „Vernetzung“ auf dezentraler Ebene d.h. zwischen den Hauptstellen eines Arbeitsamtsbezirks und seinen Nebenstellen. Arbeitsamtsbezirke sind mit ihren Nebenstellen „vernetzt“, nicht aber die verschiedenen Hauptstellen.

Der Zugang zu diesen Daten wird sehr restriktiv gehandhabt. Diese Datenbasis steht für wissenschaftliche Auswertungen nur im Rahmen von Projekten für die Arbeitsverwaltung zur Verfügung. Außerdem bedeutet die Aufbereitung dieser Daten für eine konkrete wissenschaftliche Analyse in der Regel einen erheblichen Aufwand bezüglich der Datenaufbereitung, da die personenbezogenen Informationen in verschiedenen Tabellen abgelegt sind, die in ihrer Struktur stark differieren und die Zusammenführung dieser Tabellen nicht trivial ist. Dazu kommt, dass viele Angaben in Form von Worten bzw. Abkürzungen vorliegen und diese für anspruchsvollere wissenschaftliche Auswertungen in Zahlenwerte umgewandelt werden müssen.

Trotz dieser Probleme liegen bereits wissenschaftliche Analysen zu den Effekten konkreter Maßnahmen der aktiven Arbeitsmarktpolitik auf der Basis dieser Daten vor (Almus u.a. 1999). Eine weitere Verbreitung wird diese Datenbasis allerdings nur dann finden, wenn die restriktiven Zugangsbestimmungen gelockert und die Daten der Wissenschaft in geeigneter Form zur Verfügung gestellt werden.

---

<sup>14</sup> Eine Ausnahme sind die offenen Stellen.

## **Sozialhilfestatistik**

Die Datenerhebung zur Sozialhilfe erfolgt durch deren Träger, d.h. die Kommunen, die Datenaufbereitung wird in den Statistischen Ämtern der Bundesländer durchgeführt. Dem Statistischen Bundesamt steht eine repräsentative 25% Stichprobe der in den Landesämtern verfügbaren Einzeldatensätze zur Verfügung. Diese bezieht sich nur auf Personen, die zum Stichtag 31.12. eines Jahres laufende Hilfe zum Lebensunterhalt empfangen haben, d.h. Empfänger von Hilfe in besonderen Lebenslagen werden hier nicht erfasst. Die Datei enthält Angaben zum Einkommen nach Einkommensart, zur Erwerbstätigkeit, zur Haushaltszusammensetzung und einigen anderen relevanten Variablen (z.B. Wohnkosten). Da Sozialhilfeempfänger in der EVS stark unterrepräsentiert sein dürften, stellt dies eine wichtige Datenbasis für eine gesellschaftliche Gruppe dar.<sup>15</sup> Seit kurzem besteht die Möglichkeit im Statistischen Bundesamt mit der 25%-Stichprobe der Sozialhilfestatistik wissenschaftliche Analysen durchzuführen (vgl. Burmester 2000).

### **2.1.3 Für die Wissenschaft zur Zeit nicht verfügbare Mikrodaten**

#### **Betriebsdatei der Bundesanstalt für Arbeit**

Die Betriebsdatei der BA basiert auf den Pflichtmeldungen der Arbeitgeber über deren Beschäftigte an die Sozialversicherungsträger, die von diesen an die BA weitergeleitet werden. Die Datenbasis ist somit dieselbe wie bei der oben beschriebenen IAB-Beschäftigtenstichprobe. Die BA bereitet diese Daten jeweils mit dem Stichtag zum Ende eines Quartals in der sogenannten Betriebsdatei auf. Die Betriebe werden durch die in der Datenbasis vorhandenen Betriebsnummern identifiziert. Für jeden Betrieb können somit sämtliche sozialversicherungspflichtige Beschäftigte erfasst werden. Da über die Betriebsnummer die Information zu den einzelnen Betrieben bzw. ihren Beschäftigten über die Zeit verknüpft werden kann, liegt mit der Betriebsdatei im Prinzip ein Betriebspanel als Vollerhebung der deutschen Betriebe mit sozialversicherungspflichtigen Beschäftigten vor.

Da der Betriebsdatei die gleiche Datenbasis wie der IAB-Beschäftigtenstichprobe zugrunde liegt, sind die Analysepotenziale ähnlich wie die oben bezüglich der I<sub>A</sub>BS beschriebenen. Darüber hinaus ergeben sich aufgrund des Charakters einer Vollerhebung deskriptive Auswertungsmöglichkeiten zur Erwerbstätigkeit und dem Einkommen, die mittels der I<sub>A</sub>BS nicht möglich sind. Dies betrifft beispielsweise disaggregierte regionale Analysen, aber auch Auswertungen zu bestimmten Berufsgruppen oder Wirtschaftszweigen.

Die Betriebsdatei steht bisher nur Mitarbeitern der BA zur Verfügung, externe Wissenschaftler haben nur „Ausnahmefällen“ darauf Zugriff. Nach Fritsch (1997) ist dies auf Antrag beim Bundesministerium für Arbeit- und Soziales möglich, falls Analysen aus der Arbeitsmarkt- und Berufsforschung mit anderen Daten nicht möglich wären, wobei allerdings aus Gründen des Datenschutzes nur „Auszüge“ oder „aggregierte Versionen“ zur Verfügung ge-

---

<sup>15</sup> Tabellenauswertung der Sozialhilfestatistik werden in Fachserie 13, Reihe 2 publiziert.

stellt werden. Was dies im Einzelnen bedeutet und ob es hier allgemeine Regeln gibt, entzieht sich unserer Kenntnis. Wir gehen daher davon aus, dass die Mikrodaten der Betriebsdatei für wissenschaftliche Zwecke zur Zeit nicht zur Verfügung stehen. Da die Betriebsdatei außer der Betriebsnummer, der Betriebsgröße und dem Wirtschaftszweig keine unternehmensspezifischen Angaben enthält, dürfte eine faktische Anonymisierung auch für die Betriebsdatei relativ einfach realisierbar sein. Der Wissenschaft könnte damit eine Stichprobe der Betriebe (statt der Personen wie bei der I<sub>A</sub>BS) zur Verfügung gestellt werden.

## **Kostenstrukturerhebung**

Die Kostenstrukturerhebung enthält detaillierte Angaben über die Kosten der einzelnen Produktionsfaktoren (Vorleistungen, Arbeitskosten und Zinskosten, Abschreibungen, Mieten und Pachten etc.) sowie verschiedene Produktionsgrößen (siehe Knödler 2000).<sup>16</sup> Die Kostenstrukturerhebung im Produzierenden Gewerbe wird jährlich und seit dem Berichtsjahr 1950 durchgeführt. Die Stichprobe umfaßt ca. 15 Tsd. Unternehmen mit mehr als 20 Beschäftigten und liegt zentral beim Statistischen Bundesamt vor. Das Statistische Bundesamt plant ein Anonymisierungsprojekt, das die Kostenstrukturerhebung enthält.<sup>17</sup> Seit Mitte der 90er Jahre ist eine Verknüpfung zu einem mehrjährigen Panel möglich. Die Kostenstrukturerhebung in Dienstleistungen wird alle vier Jahre erhoben und ist nicht verpflichtend.<sup>18</sup> Die Kostenstrukturstatistik im Produzierenden Gewerbe wird jährlich erhoben.

Die Mikrodaten der Kostenstrukturerhebung im Produzierenden Gewerbe wären für Analysen im Bereich Erwerbstätigkeit und Einkommen aufgrund der hohen Fallzahlen und detaillierten Erhebung der Kosten der Produktionsfaktoren für verschiedene Analysen von Interesse. Anwendungsbeispiele sind Determinanten der Arbeitsnachfrage und die Wirkungen der zunehmenden Auslagerung von Lohnarbeiten und Dienstleistungen auf die Arbeitsnachfrage der Unternehmen. Allerdings ist die Unterteilung in Angestellte und Arbeiter nicht ausreichend.<sup>19</sup>

## **Daten des Hochschul-Information-Systems**

Das HIS-*Studienberechtigtenpanel* ist eine repräsentative, bundesweite Längsschnittuntersuchung, die jeden 4. Jahrgang der Studienberechtigten untersucht. Ab 1990 werden auch die neuen Bundesländer abgedeckt. Das Studienberechtigtenpanel eignet sich zur detaillierten Analyse der Ausbildungswahl der Studienberechtigten. Das HIS-Studienberechtigtenpanel liefert Informationen über die Motive der Ausbildungswahl, unter anderem wie entscheidend

---

<sup>16</sup> Aggregierte Tabellen für Kostenarten und Produktionsmaße liegen für die vierstellige Branchenklassifikation vor und sind jeweils nach Beschäftigtengrößenklassen kreuzklassifiziert, siehe Fachserie 4 Reihe 4.3.1 – 4.3.3.

<sup>17</sup> Wir danken Herrn Sturm, Statistisches Bundesamt, für diesen Hinweis.

<sup>18</sup> Kostenstruktur in ausgewählten Dienstleistungen. Siehe Fachserie 2: Unternehmen und Arbeitsstätten, Reihe 1.1 bis 1.6.3

<sup>19</sup> Die Anzahl der Beschäftigten sind nicht für Angestellte und Arbeiter, sondern jeweils auch für Männer und Frauen verfügbar. Bruttolohn- und -gehaltsumme und Lohnnebenkosten sind jedoch nur für die ersten beiden Beschäftigtengruppen verfügbar.

Berufsziele waren. Außerdem kann der Einfluß des Arbeitslosigkeitsrisikos sowie der Einkommensaussichten, die mit den unterschiedlichen Qualifikationsniveaus und Typen (Fachrichtung) verbunden sind, auf die Studienwahl identifiziert werden. Das *HIS-Hochschulabsolventenpanel* enthält Angaben über die Betroffenheit von Arbeitslosigkeit nach dem Abschluss sowie über das monatliche Bruttoerwerbseinkommen in den einzelnen Fachrichtungen. Dies ermöglicht die Ermittlung von fachrichtungsspezifischen Arbeitslosenquoten und Erwerbseinkommen. Letztere Information liegt sonst in keiner anderen Datenbasis vor. Diese Datenbasis enthält auch detaillierte Angaben über den Studienverlauf und die Studierendauer. Letzteres Panel wird alle 4 Jahre erhoben und ermöglicht auch die Untersuchung des Übergangs von der Hochschule in das Beschäftigungssystem (siehe z. B. Lewin, Minks und Uhde 1996). Jeder Jahrgang wird zweimal befragt, einmal etwa ein Jahr nach dem Abschluß und einmal 5 Jahre später.

### **Förderstatistik der Bundesanstalt für Arbeit**

Die Bundesanstalt für Arbeit erfasst monatlich die Ein- und Austritte in bzw. aus Maßnahmen der beruflichen Weiterbildung nach Maßnahmengattung und verschiedenen Merkmalen der Teilnehmer sowie der Art der Beendigung einer bestimmten Maßnahme („Prüfung bestanden“, „vorzeitig abgebrochen“ etc.). Diese Daten werden im Verwaltungshandel der Bundesanstalt bzw. der einzelnen Arbeitsämter systematisch generiert (prozess-produzierte Daten) und weisen einen hohen Informationsgehalt auf.<sup>20</sup> Durch Verknüpfung der Mikrodaten aus der Beschäftigungsstatistik und der Förderstatistik würde für die wissenschaftliche Evaluationsforschung im Bereich der geförderten Weiterbildung eine wichtige Datenbasis zur Verfügung stehen. Im IAB wird zur Zeit an der Erstellung einer derartigen Datenbasis gearbeitet. Unseres Wissens bestehen zur Zeit aber seitens des IAB keine Überlegungen, diese Daten in faktisch anonymisierter Form der Wissenschaft allgemein zugänglich zu machen.

### **Steuerstatistik**

Durch die Änderung des Steuerstatistikgesetzes im Jahr 1996 können die Einzelangaben der verschiedenen Steuerstatistiken durch die amtliche Statistik zentral verwaltet werden. Damit liegen im Abstand von drei Jahren Einzelangaben der Lohn- und Einkommenssteuerstatistik vor. Diese Statistik ist als Totalerhebung mit knapp 30 Millionen Datensätzen die umfassendste sekundärstatistische Datenbasis der Einkommen in Deutschland. Sie erfasst jeweils für ein komplettes Veranlagungsjahr für jeden Steuerpflichtigen die Einkommen nach den verschiedenen Einkunftsarten, die Summe der Einkünfte und das zu versteuernde Einkommen sowie die Steuerschuld und Steuernachlässe. Ausserdem werden einige wenige sozio-demografische Merkmale, wie Alter, Geschlecht und Familienstand, erhoben. Diese Angaben liegen zur Zeit für rund 30 Mio. Steuerpflichtige auf Kreisebene für die Jahre 1992 und 1995 vor. Für Aus-

---

<sup>20</sup> Mit der seit 1998 gesetzlich festgelegten Verpflichtung für alle Maßnahmen der Arbeitsförderung sogenannte „Eingliederungsbilanzen“ zu erstellen, werden zukünftig entsprechende Daten auch für andere Maßnahmen (Arbeitsbeschaffungsmaßnahmen, Eingliederungshilfen etc.) systematisch erfasst werden.

wertungszwecke werden aus dieser Grundgesamtheit geschichtete Zufallsstichproben mit einer Größe von 25% und 10% gezogen (Zwick 1999).

Zur Zeit ist der Datenzugriff auf diese Statistik für externe Nutzer nur über im Statistischen Bundesamt durchgeführte Sonderauswertungen möglich. Zwick (1999) nennt einige wenige Beispiele für derartige Sonderauswertungen. Die Analyse dieser Daten durch von Wissenschaftlern entwickelte Auswertungsprogramme steht noch in der Anfangsphase. Bisher wurden auf dieser Basis unseres Wissens noch keine wissenschaftlichen Analysen durchgeführt.

#### **2.1.4 Aggregierte Daten**

Aggregierte Daten zu Erwerbstätigkeit und Einkommen liegen in vielfältiger Form vor. Gesamtwirtschaftliche und sektorale Daten zur Erwerbstätigkeit und zum Einkommen werden als Zeitreihen im Rahmen der Volkswirtschaftlichen Gesamtrechnung (VGR) nachgewiesen. Die Bundesanstalt für Arbeit publiziert laufend Daten zu Erwerbstätigkeit, Beschäftigung und Arbeitslosigkeit. Das Statistische Bundesamt veröffentlicht regelmäßig auf der Basis des Mikrozensus Kreuztabellen zur Beschäftigung nach verschiedenen Merkmalen (Formale Qualifikation, Geschlecht, Altersgruppen, Tätigkeit und Beruf).

Der Vorteil dieser aggregierten Daten besteht darin, dass sie zeitnah und für den Nutzer einfach verfügbar sind. Deren Nachteil besteht darin, dass die Analysemöglichkeiten äußerst beschränkt sind. Aus diesem Grund werden diese Daten bei wissenschaftlichen Analysen zur Erwerbstätigkeit und zum Einkommen heute kaum mehr verwendet, wie unsere Auswertung der Veröffentlichungen in diesem Bereich gezeigt hat. Deshalb werden diese Daten hier auch nur summarisch erwähnt.

#### **Volkswirtschaftlichen Gesamtrechnungen (VGR)**

Aggregierte Zeitreihendaten zu Erwerbstätigkeit und Einkommen liefert die Volkswirtschaftliche Gesamtrechnung (VGR). Diese enthält auch Angaben über die insgesamt geleisteten Arbeitsstunden und Angaben über die Arbeitskosten einschliesslich Lohnnebenkosten. Die Volkswirtschaftliche Gesamtrechnung dient als Basis für eine Vielzahl empirischer Untersuchungen auf dem Gebiet des Arbeitsmarktes. Der Vorteil der Erwerbstätigenzahlen auf Basis der VGR ist die zeitnahe Erhebung. Von Nachteil ist die relativ schlechte Datenqualität und häufige Revidierung am aktuellen Rand. Dies ergibt sich daraus, dass die Erwerbstätigenzahlen der VGR im wesentlichen durch Fortschreibung der letzten Volks- und Arbeitsstättenzählung unter hauptsächlichlicher Verwendung des Mikrozensus und der Beschäftigtenstatistik gewonnen werden (siehe hierzu ausführlich Schares 1998).

Gesamtwirtschaftliche Daten zum Einkommen sind aus der Einkommensrechnung der VGR verfügbar. Daten zur Verteilung der Einkommen nach Haushaltsgruppen werden vom Statistischen Bundesamt auf der Basis der VGR und anderer Datenquellen (Mikrozensus, Einkommens- und Verbrauchsstichprobe, Lohn- und Einkommenssteuerstatistik) im Rahmen von Sonderveröffentlichungen publiziert (bspw. Fachserie 15, Wirtschaftsrechnungen. Reihe 1, Einnahmen und Ausgaben ausgewählter privater Haushalte, Schüler 1987, sowie Essig

2000 zum System der Europäischen VGR). Die Daten sind in gruppierter Form (nach Erwerbsstatus, Haushaltstyp und Anzahl der Kinder) verfügbar.

Durch die Umstellung der nationalen VGR auf das Europäische System der Volkswirtschaftlichen Gesamtrechnung (ESVG) im Jahr 1999 liegen konsistente VGR-Daten (auf Jahres- und Quartalsbasis) für Deutschland nur bis 1991 vor. Die VGR in Deutschland wird seit 1999 nur mehr auf der Basis des ESVG fortgeführt. Auch gibt es keine Unterscheidung zwischen Ost- und Westdeutschland. Gegenüber der früheren deutschen VGR weist die auf Basis des ESVG erstellte VGR ein wesentlich höheres Beschäftigungsniveau (ca. 2 Millionen Beschäftigte) auf.

Das Deutsche Institut für Wirtschaftsforschung (DIW) erstellt in Absprache mit dem Statistischen Bundesamt ebenfalls eine vierteljährliche Volkswirtschaftliche Gesamtrechnung. Das DIW weist im Vergleich zum Statistischen Bundesamt eine weitere Entstehungsrechnung aus, die auf dem Arbeitsvolumen (tägliche Arbeitszeit  $\times$  Arbeitstage  $\times$  im Inland Erwerbstätige) beruht. Auch die VGR des DIW wird nach dem ESVG fortgeführt.

Ähnliche Berechnungen zur Verteilung der Einkommen nach Haushaltsgruppen wie vom Statistischen Bundesamt werden auch vom DIW mehr oder weniger regelmäßig publiziert. Auch diese Berechnungen basieren auf Zusammenführungen von Daten aus der VGR, der Bevölkerungsstatistik sowie von Einkommensstrukturdaten, wie bspw. den Einkommens- und Verbrauchsstichproben. Selbstverständlich sind auch diese Daten nur nach groben Gruppen und nicht als Individualdaten gegeben. Für die einzelnen Gruppen sind jeweils die Gruppengrenzen, die Besetzungszahlen für die einzelnen Gruppen und die Mittelwerte der Einkommen in den einzelnen Gruppen gegeben.

### **Aggregierte Daten der Bundesanstalt für Arbeit**

Die Bundesanstalt für Arbeit publiziert monatlich in deren Amtlichen Nachrichten über Arbeitslose und offene Stellen (sowohl Bestände als auch Stromgrößen), unterteilt nach verschiedenen Merkmalen (Alter, Geschlecht, Bundesland). Die Daten beziehen sich auf die bei den Arbeitsämtern registrierten Arbeitslosen und gemeldeten offenen Stellen. Diese Daten sind auch für Arbeitsamtsbezirke verfügbar.

Darüber hinaus werden die Bezieher von Leistungen der Arbeitslosenversicherung und die Anzahl der Teilnehmer an Maßnahmen der Arbeitsförderung (aktive Arbeitsmarktpolitik) auf Monatsbasis und nach Arbeitsamtsbezirken erhoben. Diese Daten können kostenpflichtig von der Bundesanstalt auch in tabellarischer Form auf Datenträgern bezogen werden. Für das Jahr 1998 wurden von der Bundesanstalt auch Daten zu den sogenannten Eingliederungsbilanzen publiziert. Diese Eingliederungsbilanzen enthalten auf der Ebene der Arbeitsamtsbezirke u.a. Informationen über den Anteil der Teilnehmer an den einzelnen Maßnahmen, die sechs Monate nach Beendigung einer Maßnahme nicht mehr als Arbeitslose registriert sind.

Mit der Beschäftigtenstatistik verfügt die Bundesanstalt über eine Vollerhebung aller sozialversicherungspflichtigen Beschäftigten in Deutschland. Diese Statistik basiert auf Meldungen der Betriebe an die Sozialversicherungsträger, welche die Daten an die Bundesanstalt weiterleiten. Die Verteilung der sozialversicherungspflichtigen Beschäftigten nach verschiedenen Merkmalen (u.a. Wirtschaftszweig, ausgeübter Beruf, Nationalität etc.) wird zu bestimmten Stichtagen ebenfalls in den Amtlichen Nachrichten der Bundesanstalt ausgewiesen.

In der einmal pro Jahr durchgeführten sogenannten „Strukturerhebung“ werden detaillierte Angaben zu Beständen sowie Zu- und Abgängen an Arbeitslosen und offenen Stellen ausgewiesen. Die Arbeitslosen werden nach bestimmten Merkmalen (Alter, bisherige Dauer der Arbeitslosigkeit, Berufsausbildung und Berufsbereichen, gesundheitliche Einschränkungen) differenziert und untergliedert nach Arbeitsamtsbezirken ausgewiesen.

### **Tabellenauswertungen des Mikrozensus**

Das Statistische Bundesamt publiziert regelmäßig Statistiken zur Erwerbstätigkeit auf der Basis des Mikrozensus in Form umfangreicher Tabellen.<sup>21</sup> Die Anzahl der Erwerbstätigen ist nach formaler Qualifikation, Geschlecht, Altersgruppe, Tätigkeit, beruflicher und betrieblicher Stellung klassifiziert. Nur wenige Variablen sind kreuzklassifiziert (beispielsweise formale Qualifikation und Altersgruppen jeweils für Männer und Frauen, Beruf und Arbeitszeit, Beruf und Stellung im Beruf, Beruf und Branche). Einkommen werden nur nach (groben) Einkommensklassen erhoben. Ein Nachweis der auf die verschiedenen Einkommensarten entfallenen Einkommenshöhe erfolgt im Mikrozensus nicht. Nachgewiesen wird nur die primäre Einkommensart und ob darüber hinaus Einkommen aus den einzelnen anderen erfassten Einkommensarten bezogen wurde. Aufgrund dieser Beschränkungen sind die Einkommensangaben im Mikrozensus für Verteilungsanalysen wenig geeignet.

### **Tabellenauswertungen der Einkommens- und Verbrauchsstichprobe (EVS)**

Detaillierte Auswertungen zum Einkommen nach Haushaltstypen auf der Basis der oben beschriebenen EVS werden in Tabellenform vom Statistischen Bundesamt in Einzelveröffentlichungen in der Fachserie 15 publiziert. Es werden Durchschnittseinkommen für einzelne Haushaltstypen und die Zusammensetzung der Einkommen detailliert nachgewiesen. Die in der Fachserie publizierten Tabellen ermöglichen zwar die Darstellung der durchschnittlichen Einkommen nach einzelnen Haushaltsgruppen und deren zeitlichen Entwicklung, detailliertere Analysen sind damit jedoch nicht möglich. Wie oben beschrieben stellt das Statistische Bundesamt zu diesem Zweck seit kurzem die Individualdaten der EVS in faktisch anonymisierter Form zur Verfügung.

### **Tabellenauswertungen der laufenden Wirtschaftsrechnungen**

Eine weitere Datenquelle für Einkommen sind die laufenden Wirtschaftsrechnungen, die hinsichtlich der Definition und Erfassung der Einkommen der EVS entsprechen.<sup>22</sup> Einbezogen werden aber nur Haushalte von drei ausgewählten Haushaltsgruppen, die aber nur einen relativ geringen Teil aller Haushalte repräsentieren. Diese Datenbasis erlaubt daher keine Aussa-

---

<sup>21</sup> Diese Statistiken sind in den Fachserie 1, Reihe 4.1.1. „Stand und Entwicklung der Erwerbstätigkeit“ und Reihe 4.1.2. „Bevölkerung und Erwerbstätigkeit“ zusammengefasst.

<sup>22</sup> Siehe hierzu Statistisches Bundesamt, Fachserie 15, Wirtschaftsrechnungen. Reihe 1, Einnahmen und Ausgaben ausgewählter privater Haushalte..

gen über das Durchschnittseinkommen oder gar die Verteilung des Einkommens in der Gesamtpopulation.

## 2.2 Defizite und Datenbedarf

Die Daten des Statistischen Bundesamts und der Bundesanstalt für Arbeit befriedigen nur zum Teil die Analyse- und Informationsbedürfnisse der Wissenschaft. Generell gilt aus wissenschaftlicher Sicht auch für den Bereich Erwerbstätigkeit, dass die Amtliche Statistik einschliesslich der BA einerseits zu viele Daten erfasst, die relativ geringe Relevanz haben, andererseits aber Daten und Informationen im Rahmen behördlicher Arbeit existieren, die für wissenschaftliche Untersuchungen sehr interessant wären, aber zur Zeit nicht oder nur sehr eingeschränkt verfügbar sind. Zudem sind Vollerhebungen nicht immer notwendig (siehe auch Hahlen 1999). Im Verarbeitenden Gewerbe wäre eine Stichprobe wie die Kostenstrukturerhebung für wissenschaftliche Zwecke vollkommen ausreichend. Die fehlende Dienstleistungsstatistik führt im Bereich Erwerbstätigkeit zu einem großen Informationsdefizit. Alternative Datenquellen wie die Beschäftigtenstatistik umfassen nur sozialversicherungspflichtig Beschäftigte. Jedoch ist der Deckungsgrad der Beschäftigtenstatistik gerade in den neuen Dienstleistungen wegen des hohen Selbständigenanteils besonders niedrig. Der Mangel an aussagekräftigen Daten auf dem Gebiet der Dienstleistungen hat in den letzten Jahren zu einem erheblichen Anstieg der Datenproduktion geführt. Beispielsweise findet seit 1994 die ZEW/Creditreform Konjunkturumfrage bei unternehmensnahen Dienstleistern statt (Kaiser 1999).

### 2.2.1 Defizite verfügbarer Mikrodaten

Obwohl mittlerweile mehrere Mikrodatsätze zur Erwerbstätigkeit und zum Einkommen für wissenschaftliche Zwecke frei verfügbar sind, weisen diese doch alle mehr oder weniger gravierende Defizite auf, welche die Analysemöglichkeiten erheblich einschränken:

- Das **Sozio-ökonomische Panel** enthält sowohl detaillierte Angaben zur Erwerbstätigkeit als auch zum Einkommen. Der Nachteil des SOEP gegenüber den im folgenden erwähnten Mikrodatsätzen der amtlichen Statistik besteht im geringen Stichprobenumfang. Deskriptive Analysen für kleine Gruppen mit Hochrechnungen sind daher kaum oder nur sehr eingeschränkt möglich. Ausserdem liegt auch beim SOEP – wie bei den meisten Panelstudien – eine hohe Panelsterblichkeit vor. Aufgrund der Freiwilligkeit an der Teilnahme tritt bei einigen Variablen auch eine große Anzahl fehlender Werte auf. Bestimmte Einkommensarten, wie Einkünfte aus Unternehmertätigkeit, Vermietung und Verpachtung und Kapitalvermögen, sind im SOEP nicht oder völlig unzureichend erfasst. Die geringen Fallzahlen erlauben keine oder nur sehr eingeschränkte Analysen zu bestimmten Fragestellungen (bspw. der Wirkungsweise arbeitsmarktpolitischer Maßnahmen) oder bestimmter Gruppen (bspw. erwerbstätiger Sozialhilfeempfänger).

- Die relativ geringen Fallzahlen und die beschränkte Information zur Erwerbstätigkeit und zum Einkommen sind auch die entscheidenden Nachteile der **ALLBUS**-Befragungen, deren Schwerpunkte allerdings im sozialwissenschaftlichen Bereich liegen.
- Das **Europäische Haushaltspanel** (ECHP) enthält ebenfalls Angaben sowohl zur Erwerbstätigkeit als auch zum Einkommen und bietet darüber hinaus das Potenzial für vergleichende Analysen. Das größte Defizit des ECHP in der Version, die dem Nutzer von EUROSTAT zur Verfügung gestellt wird („User Data Base“) besteht in der starken Aggregation von Werten bei wichtigen Variablen (z.B. Ausbildung), der fehlenden Information zum individuellen Bruttoerwerbseinkommen und generell der Behandlung fehlender Werte („missings“). Darüber hinaus wird der Zugang zu den Daten für die anderen Länder von EUROSTAT restriktiv gehandhabt, was auch durch relativ hohe Preise zum Ausdruck kommt.
- Der **Mikrozensus** enthält zwar detaillierte Informationen zur Erwerbstätigkeit, aber keine Angaben zum Erwerbseinkommen. Aufgrund der unzureichenden Erfassung des Einkommens im MZ (fehlende Differenzierung nach Einkunftsarten, Erfassung des Einkommens in (groben) Klassen) eignet sich der MZ nicht für Einkommensanalysen. Das Analysepotenzial des MZ für Querschnittsanalysen zur Erwerbstätigkeit ist durch das Fehlen von Angaben zum Erwerbseinkommen ebenfalls stark eingeschränkt. Eine weitere Beschränkung des MZ besteht darin, dass aufgrund der fehlenden Panelkomponente Verlaufsanalysen nur sehr eingeschränkt möglich sind.
- Die **Einkommens- und Verbrauchsstichprobe** enthält zwar detaillierte und relativ genaue Angaben zum Einkommen, aber keine Information zum Erwerbsumfang, insbesondere auch keine Angaben zu den geleisteten und/oder bezahlten Arbeitsstunden. Damit sind zwar detaillierte Analysen zur Einkommensstruktur möglich, die wichtigen Zusammenhänge zwischen Einkommen und Erwerbstätigkeit können aber nicht oder nur sehr eingeschränkt analysiert werden. Eine Beschränkung der EVS besteht darin, dass die EVS bisher nur alle fünf Jahre durchgeführt wird. Ein weiterer Nachteil der EVS besteht darin, dass sie als reine Querschnitterhebung angelegt ist und daher Einkommensanalysen im Lebenszyklus nur sehr beschränkt möglich sind.
- Die **laufenden Wirtschaftsrechnungen** decken nur einen kleinen Teil aller Haushalte ab und enthalten im Vergleich zur EVS wenig Information. Sie eignen sich daher weder zur Fortschreibung der aus der EVS gewonnenen Zahlen noch für wissenschaftliche Analysen zum Konsumentenverhalten.
- Die **Zeitbudgeterhebung 1991/92** als eine Spezialerhebung zur Erwerbstätigkeit auch ausserhalb des Marktsektors weist vor allem das Defizit auf, dass das individuelle Erwerbseinkommen nicht erhoben wird. Darüber hinaus sind keine retrospektiven Angaben zur Erwerbsbiographie vorhanden. Beides schränkt die Analysemöglichkeiten bezüglich Fragestellungen, für die sich diese Erhebung im Prinzip eignen würde, erheblich ein.
- Die **IAB-Beschäftigtenstichprobe** (I<sub>A</sub>BS) enthält Angaben zur Erwerbstätigkeit und zum Erwerbseinkommen (Entgelte) nach wichtigen Merkmalen der sozialversicherungspflichtig Beschäftigten. Registrierte Arbeitslosigkeit wird im Rahmen dieser Statistik ebenfalls erfasst. Ein wesentlicher Nachteil der I<sub>A</sub>BS besteht darin, dass die Entgelte bei der oberen Sozialversicherungsgrenze abgeschnitten (zensiert) sind. Dies trifft auf den Großteil der beschäftigten Hochschulabsolventen zu. Ein weiterer wesentlicher Nachteil dieser Datenbasis besteht darin, dass die Anzahl der geleisteten Arbeitsstunden nicht erfasst wird. Ein

Nachteil der I<sub>A</sub>BS ist auch, dass die zwei bisher erstellten Stichproben für die Wissenschaft erst mit großer zeitlicher Verzögerung verfügbar wurden, wodurch das Interesse von Wissenschaftlern damit zu arbeiten vermutlich geringer war als dies bei einer zügigen Bereitstellung der Daten der Fall gewesen wäre.

- Die **BIBB/IAB Erhebung „Qualifikation und Berufsverlauf“** enthält detaillierte Informationen zur Qualifikationsstruktur und Erwerbstätigkeit im Abstand von fünf Jahren. Eine wesentliche Beschränkung dieser Datenbasis besteht darin, dass sie erst nach Ablauf einer relativ langen Zeit an interessierte Wissenschaftler weitergegeben wird. Dies ist vermutlich ein wesentlicher Grund dafür, dass diese Datenbasis durch die Wissenschaft bisher kaum genutzt wurde.
- Ein Nachteil der **Gehalts- und Strukturhebung (GLS)** für Analysezwecke besteht darin, dass diese wichtige Datenbasis nicht beim Statistischen Bundesamt, sondern für die einzelnen Bundesländer bei den jeweiligen Statistischen Ämtern der Länder verfügbar ist. Ein weiterer Nachteil dieser Datenbasis besteht darin, dass sie nur alle fünf Jahre erhoben wird. Ein wesentlicher Nachteil dieser Datenbasis besteht auch in der Beschränkung auf das Produzierende Gewerbe, den Handel sowie Banken und Versicherungen, womit die zunehmende Bedeutung anderer unternehmensbezogener und persönlicher Dienstleistungen nicht erfasst wird. Dies gilt in noch stärkerem Ausmaß für die **Amtliche Industriestatistik**, die nur das Produzierende Gewerbe erfasst.
- Das **IAB-Betriebspanel** erfasst hingegen alle Betriebe mit sozialversicherungspflichtigen Beschäftigten. Das Analysepotenzial dieser wichtigen Datenbasis für die Wissenschaft ist jedoch stark eingeschränkt, da die Zugangsmöglichkeiten bisher zu restriktiv gehandhabt werden. Die beim IAB eingerichtete und durch das BMBF finanzierte Schalterstelle, über die von externen Wissenschaftlern verfasste Auswertungsprogramme bearbeitet werden sollen, erscheint für komplexere mikroökonomische Analysen ungeeignet.

Neben den Beschränkungen des Analysepotenzials, die sich aus den aufgezeigten Defiziten der frei oder beschränkt zugänglichen Mikrodatenbasen ergeben, enthalten die für Deutschland verfügbaren Mikrodaten zu bestimmten Fragestellungen zu den Bereichen Erwerbstätigkeit und Einkommen keine oder nur unzureichende Informationen. Dies betrifft unter anderem die folgenden Bereiche, die international im Zentrum der wissenschaftlichen Forschung stehen, aber auch von erheblichem wirtschaftspolitischem Interesse sind:

- Betriebs- oder unternehmensbezogene Mikrodaten sind bisher nicht oder nur sehr eingeschränkt verfügbar. Dies liegt zum einen in der Problematik der Bereitstellung dieser Daten für wissenschaftliche Zwecke in einer Form, die zum einen den Datenschutzbestimmungen genügt (siehe dazu Abschnitt 6), andererseits das Analysepotenzial derartiger Daten gewährleistet.
- Die einzelnen Erhebungen, die jeweils nur partiell die für viele Analysen erforderliche Information enthalten, stehen unverbunden nebeneinander, die Verknüpfung der relevanten Information aus zwei oder mehreren Datensätzen ist nicht oder nur über mit großen Unsicherheiten behafteten statistischen Verfahren möglich. Im Haushaltsbereich betrifft dies vor allem den Mikrozensus und die Einkommens- und Verbrauchsstichprobe. Der MZ enthält detaillierte Angaben zur Erwerbstätigkeit, aber keine Angaben zur Höhe des Erwerbseinkommens. Dieses wird in der EVS zwar sehr genau erfasst, diese Datenbasis enthält aber kaum Angaben zur Erwerbstätigkeit. Für ökonomische Analysen zur Er-

werbstätigkeit ist aber das Erwerbseinkommen (bezogen auf ein bestimmtes Stundenvolumen) eine zentrale Variable. Im Unternehmensbereich betrifft dies insbesondere das Nebeneinander der Gehalts- und Lohnstrukturerhebung (GLS) und der Amtlichen Industriestatistik.

- Die wichtigsten Datensätze der Amtlichen Statistik (Mikrozensus, Einkommens- und Verbrauchsstichprobe, Gehalts- und Lohnstrukturerhebung) sind als reine Querschnitterhebungen angelegt. Damit entsprechen sie nur eingeschränkt den Anforderungen der wissenschaftlichen Forschung, die sich zunehmend auf Analysen von Paneldaten (Daten mit wiederholten Beobachtungen der gleichen Untersuchungseinheiten) beziehen.
- Generell besteht ein Mangel an unterjährigen Mikrodaten, insbesondere bei personen- und haushaltsbezogenen Erhebungen. Dies betrifft vor allem auch den Mikrozensus als die wichtigste allgemeine Erhebung zur Erwerbstätigkeit in Deutschland. Im Statistischen Bundesamt bestehen zur Zeit Überlegungen der Ausweitung des MZ auf eine unterjährige Erhebung. Dies würde auch die Möglichkeit eröffnen, nach dem Rotationsprinzip Personen innerhalb eines kürzeren Zeitraums mehrmals zu befragen. Damit stünde beschränkt auch Panelinformation im Rahmen des Mikrozensus zur Verfügung.
- Für detaillierte Analysen zu Erwerbstätigkeit und Einkommen von Hochqualifizierten stehen nur sehr eingeschränkt Informationen zur Verfügung. Das SOEP erlaubt aufgrund der relativ kleinen Stichprobe keine detaillierten Analysen bspw. nach Studienabschlüssen. Im MZ wird diese Information seit der Änderung im Jahr 1996 nur noch alle vier Jahre erhoben. Darüber hinaus sind die Analysemöglichkeiten des MZ in diesem Bereich dadurch stark beschränkt, dass keine Angaben zum Erwerbseinkommen erhoben werden. Eine informative Datenbasis, die auch Panelanalysen ermöglichen würde, wird aus öffentlichen Mitteln seit Jahren erhoben und von HIS aufbereitet, steht externen Wissenschaftlern jedoch nicht zur Verfügung.
- Mikrodaten mit kleinräumiger Regionalinformation (Kreise, Arbeitsamtsbezirke) sind für wissenschaftliche Analysen nicht oder nur in Ausnahmefällen verfügbar. Diese Information wird zwar in der Regel erhoben, aber aus Datenschutzgründen nicht weitergegeben. Dies betrifft bspw. den Mikrozensus und die I<sub>A</sub>BS sowie das SOEP. Daher sind Analysen zur regionalen Mobilität für Deutschland nicht oder nur sehr eingeschränkt möglich. Aufgrund der großen Bedeutung der regionalen Mobilität zum Arbeitsmarktausgleich klafft hier zur Zeit eine große Forschungslücke.
- Die Datenlage zur Erwerbstätigkeit, die nicht dem „Standard-Arbeitsverhältnis“ entspricht, ist sehr dürftig. Selbständigkeit und Scheinselbständigkeit, befristete Arbeitsverhältnisse, geringfügige Beschäftigung und Leiharbeit werden immer wichtiger, insbesondere im Bereich neuerer Dienstleistungen. In der Amtlichen Statistik werden diese neuen Formen der Erwerbstätigkeit nicht oder nur unzureichend erfasst. Beispielhaft dafür sei die Unsicherheit über das Ausmaß der geringfügigen Beschäftigung genannt. Während der Mikrozensus als die wichtigste Datenbasis zur Erwerbstätigkeit das Ausmaß der geringfügigen Beschäftigung zuletzt mit ca. 2 Millionen Personen ausgewiesen hat, wurden auf der Basis kleiner Stichproben hochgerechnete Zahlen von 5 – 6 Millionen berechnet. Diese mit großen Unsicherheiten behafteten Zahlen wurden als Basis für weitreichende sozialrechtliche Änderungen herangezogen, die Zahlen des Mikrozensus wurden von Wissenschaft und Politik nicht zur Kenntnis genommen.

- Mikrodaten zur Teilnahme an durch die Bundesanstalt für Arbeit geförderten Maßnahmen der „aktiven“ Arbeitsmarktpolitik sind zwar als prozessproduzierte Daten vorhanden und liegen für den Bereich der geförderten beruflichen Weiterbildung auch in Form der Förderstatistik bei der BA vor. Diese Datenbasis, die auch mit der Beschäftigtenstatistik der BA („Betriebsdatei“) verknüpft werden könnte, ist aber für Forschungszwecke bisher nicht zugänglich.
- Die einzelnen Datensätze der amtlichen Statistik basieren häufig auf unterschiedlichen Klassifikationen (beispielsweise von Qualifikationen und Wirtschaftszweigen) und Abgrenzungen, die häufig untereinander inkompatibel sind und somit die Verknüpfung der Information aus unterschiedlichen Statistiken erschweren oder unmöglich machen. Auch sind die verwendeten Klassifikationsschemata häufig ziemlich veraltet. Beispielsweise haben sich in den letzten Jahren eine Reihe von neuen Berufen im Informationstechnologie- und Multimediabereich entwickelt, die sich in der amtlichen Statistik nicht wiederfinden.

Generell gilt, dass aufgrund des aufgezeigten Mangels an informativen und frei verfügbaren Daten zu einzelnen Themenschwerpunkten in den Bereichen Erwerbstätigkeit und Einkommen von wissenschaftlichen Instituten und einzelnen Wissenschaftlern Daten mit meist sehr kleinen Stichproben erhoben werden. Beispielsweise werden im Auftrag der BA bzw. dem IAB häufig Personen- und Firmenbefragungen mit verschiedenen Schwerpunkten im Bereich Erwerbstätigkeit durchgeführt, die der Wissenschaft allerdings meist für Sekundäranalysen nicht zur Verfügung stehen. Von der DFG wurde die Anlauffinanzierung für das Firmenpanel Niedersachsen übernommen. Das BIBB befragt regelmäßig Betriebe zur Ausbildungsbeteiligung.<sup>23</sup> So verdienstvoll und aufgrund der gegebenen Datenlage notwendig diese Befragungen auch sein mögen, sie können keinen Ersatz für eine breite, nach einheitlichen Konzepten und kontinuierlich erhobene Datenbasis sowohl für Personen und Haushalte als auch Betriebe und Unternehmen darstellen. Darüber wird durch die weiter zunehmende Befragungsintensität bei Haushalten und Unternehmen die Antwortbereitschaft massiv beeinflusst, worunter die Rücklaufquoten und die Datenqualität in erheblichem Ausmaß leiden.

### 2.2.2 Defizite von aggregierten Daten

Die amtlichen Statistiken zu Erwerbstätigkeit und Einkommen weisen eine Reihe von Lücken auf. Die fehlende Dienstleistungsstatistik ist die bedeutsamste Lücke im wirtschaftsstatistischen System Deutschlands. Das Statistische Bundesamt, Ministerien und Wirtschaftsverbände fordern seit Jahren die Einführung einer Dienstleistungsstatistik. Ein sogenanntes Dienstleistungsstatistikgesetz war in der Vergangenheit vorgesehen, ist jedoch aufgrund politischen Widerstandes bis heute nicht verabschiedet worden. Eine Dienstleistungsstatistik würde die Basis für eine zeitnahe Erhebungen auch in der IK Branche schaffen, beispielsweise zu Beschäftigung und Arbeitskosten. Die geeigneten Statistik-Methoden sind durch Testerhebungen des Statistischen Bundesamtes seit langem geklärt. In der Zwischenzeit liegt ein Gesetzesentwurf vor und soll in dieser Wahlperiode verabschiedet werden. Eine Folge dieses Defizits ist,

<sup>23</sup> Die im Anhang erwähnte BIBB-Befragung Referenz-Betriebs-System (RBS) umfasst derzeit rund 1.700 Ausbildungsbetriebe und Institutionen.

dass unabhängige Wirtschaftsforschungsinstitute in diesem Bereich Daten erheben lassen.<sup>24</sup> Grösstenteils wird auch auf die Beschäftigtenstatistik oder den Mikrozensus zurückgegriffen.

Neben der fehlenden Dienstleistungsstatistik gibt es eine Reihe von weiteren Schwachstellen der Amtlichen Statistiken zu Erwerbstätigkeit und Einkommen auf der Basis aggregierter Daten:

- Die in der **VGR** nachgewiesenen aggregierten Zahlen zur Erwerbstätigkeit und Einkommen sind durch die Umstellung der VGR auf das Europäische System der Volkswirtschaftlichen Gesamtrechnungen nur mehr ab 1991 nach einem einheitlichen Erhebungsschema verfügbar. Bereits durch die Wiedervereinigung wurden wichtige Zeitreihen für das frühere Bundesgebiet getrennt nur bis 1994 erhoben. Die zur Zeit vorliegenden aggregierten Zeitreihen der VGR sind für wissenschaftliche Analysen zu kurz. Außerdem war die Umstellung der VGR mit einer deutlichen Revision der Beschäftigtenzahlen verbunden (diese wurden um ca. 2 Mio. nach oben revidiert). Die geringfügige Beschäftigung wird in der Amtlichen Statistik auf der Basis des Mikrozensus berechnet und damit nach allgemeiner Auffassung deutlich unterschätzt. Auch die temporäre Ausländerbeschäftigung dürfte aufgrund der Untererfassung bestimmter Sonderformen wie Saisonarbeiter, Werkvertragsbeschäftigte sowie Grenz- und Gastarbeitnehmer nicht ihrer tatsächlichen Bedeutung gemäß erfasst sein. Darüber hinaus weist die VGR zeitnah nur zusammengefasste Wirtschaftsbereiche aus. Über die sozialversicherungspflichtige Beschäftigung wird mit einer zeitlichen Verzögerung von neun Monaten informiert und dann auch nur in nicht ausreichend tiefer Branchengliederung.
- Die „**Strukturerhebung**“ der **BA** ist zwar bezüglich der Struktur der Arbeitslosigkeit relativ detailliert, wird aber nur einmal pro Jahr durchgeführt. Gerade was die Arbeitslosigkeit als (durchschnittlich) relativ kurzfristiges Phänomen anbelangt, wären mehrmalige Auswertungen der Grunddaten wünschenswert. Entsprechend detaillierte unterwertige Auswertungen zur (sozialversicherungspflichtigen) Beschäftigung im Rahmen der Strukturerhebungen wären ebenfalls wünschenswert.
- Die publizierten **Tabellenauswertungen des MZ** zu Bevölkerung und Erwerbstätigkeit erlauben keine tiefer gehenden Analysen, da die Tabellenbände überwiegend bivariate Kreuztabellen enthalten. Ein entscheidender Nachteil der Tabellen ist, dass Erwerbstätige ohne Berufsabschluss nicht separat ausgewiesen sind, sondern implizit mit den Nichtantwortern zu einer Restgruppe zusammengefasst sind. Die Qualifikatorische Arbeitslosenquote für Ungelernte kann somit nur mit den Einzeldaten berechnet werden. Die publizierten **Tabellenauswertungen der EVS** weisen zwar die Durchschnittseinkommen nach Haushaltsgruppen nach, erlauben aber keine detaillierteren Verteilungsanalysen.
- In der **Amtlichen Industriestatistik**, die sich durch vergleichsweise hohe Aktualität auszeichnet, werden nur Unternehmen mit mehr als 20 Beschäftigten beobachtet. Ein zeitnahes Erfassen der für die gesamtwirtschaftliche Beschäftigung bedeutsamen Kleinbetriebe ist nicht möglich.
- Es liegen zwar Zeitreihen zur Beschäftigung (sozialversicherungspflichtig Beschäftigte) und Arbeitslosigkeit auf der Ebene von Kreisen vor, die Verfügbarkeit derartiger Statistiken ist aber nur beschränkt gegeben. Einige Statistische Landesämter bieten auf ihrer

---

<sup>24</sup> Seit 1994 führt das ZEW gemeinsam mit der Creditreform eine Konjunkturumfrage in zehn unternehmensnahen Dienstleistungsbranchen durch (siehe Kaiser 1999 und die Projektbeschreibung im internet unter [http://www.zew.de/projekte/industrieoekonomik/laufende\\_p/io18.html](http://www.zew.de/projekte/industrieoekonomik/laufende_p/io18.html))

Webseite Beschäftigungs- und Arbeitslosigkeitsdaten auf der Ebene von Kreisen an. Beispielsweise weist das Statistische Landesamt Baden-Württemberg sozialversicherungspflichtig Beschäftigte und Arbeitslose für verschiedene Merkmale auf der Ebene von Kreisen an.<sup>25</sup>

Generell gilt, dass – bedingt durch die Wiedervereinigung – Zeitreihen von Statistiken zu Erwerbstätigkeit und Einkommen oft sehr kurz und für empirische Analysen wenig geeignet sind. Hinzu kommt die Umstellung auf die NACE-Klassifikation. Das Kernproblem tabellarischer Datenanalysen ist zudem die Möglichkeit von Scheinkorrelationen. Mehrdimensionale Kreuztabellen können dieses Problem zwar reduzieren, deren Interpretation wird mit steigender Dimension aber immer schwieriger.

### 2.3 Vernetzungsmöglichkeiten

Wie bereits bei der Darstellung der Defizite die Datenlage betreffend erwähnt wurde, sind für wissenschaftliche Analysen zu Erwerbstätigkeit und Einkommen die notwendigen Informationen in einer einzelnen Erhebung nicht vorhanden. Als Defizit wurde in diesem Zusammenhang das weitgehend unverbundene Nebeneinander der einzelnen Erhebungen der Amtlichen Statistik identifiziert. Als Beispiele dafür wurden der Mikrozensus und die Einkommens- und Verbrauchsstichprobe genannt. Beispiele im Bereich der Wirtschaftsstatistik sind die Amtliche Industriestatistik und die Kostenstrukturerhebung. Durch Vernetzung der einzelnen Erhebungen bzw. durch Verknüpfung relevanter Information könnte das Analysepotenzial der Amtlichen Statistik für die Wissenschaft wesentlich erhöht werden.

Eine Möglichkeit der stärkeren inhaltlichen Vernetzung der Amtlichen Statistik besteht im Konzept des sogenannten „Access Panels“, wie dies jüngst auch von einer beim Statistischen Bundesamt angesiedelten Arbeitsgruppe für den Bereich der Haushaltserhebungen vorgeschlagen wurde (Bechtold und Chlumsky 2000). Die Grundidee dieses für Deutschland neuen Konzepts besteht darin, Informationen zu unterschiedlichen Bereichen aus verschiedenen Haushaltsstatistiken sowohl im Quer- als auch im Längsschnitt auf der Basis eines einheitlichen Stichprobenkonzepts in einer Form darzustellen. Dabei könnte der Mikrozensus im Bereich der Haushaltsstatistiken die Basiserhebung bilden, aus der das Access-Panel aufgebaut werden soll, das für die übrigen („vernetzten“) Haushaltserhebungen die Ausgangsdatei („Master Sample“) bildet. Damit könnte bspw. die Einkommens- und Verbrauchsstichprobe in Form einer jährlichen Erhebung, die Zeitbudgeterhebung und andere Zusatzbefragungen mit dem Mikrozensus vernetzt und damit auf ein einheitliches Erhebungskonzept umgestellt werden. Im Prinzip ließe sich dieses Konzept auch auf betriebs- und unternehmensbezogene Daten der Amtlichen Statistik übertragen.

Die wesentlichen Vorteile dieses Konzepts, das im übrigen in der privaten Umfrageforschung nach Bechtold und Chlumsky (2000, S. 83) schon seit über 50 Jahren erfolgreich eingesetzt wird, sind: Zum einen kann für die einzelnen Zusatzerhebungen auf eine Zufallsstichprobe mit fast vollständiger Teilnahme und bekannten Merkmalen der Teilnehmenden zurückgegriffen werden (hier: den Mikrozensus), wodurch Hochrechnungen vereinfacht werden. Zum anderen müssen bei der Zusatzerhebung viele Merkmale, die bei einer isolierten Befragung unabdingbar sind, nicht erhoben werden, da diese bereits aus der Basiserhebung bekannt

---

<sup>25</sup> Siehe <http://www.statistik.baden-wuerttemberg.de/SRDB/>

sind. So könnte man sich beispielsweise bei einer integrierten Einkommens- und Verbrauchsstichprobe auf die detaillierte Erfassung der Einkommen (und Ausgaben) beschränken und detaillierte Angaben zur Erwerbstätigkeit und anderen wichtigen Merkmalen aus dem Mikrozensus übernehmen.

Vernetzungsmöglichkeiten bestehen auch bezüglich der Zusammenführung von Firmendaten mit Personendaten. Versuche dazu gibt es bereits im Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit. Dabei werden die Betriebsdaten des IAB-Betriebspanels über die Betriebsnummer mit der Beschäftigtenstatistik verknüpft (vgl. Bellmann et al. 1999). Damit wird ein sehr informativer Datensatz geschaffen, da für jedes Jahr ca. 10 Millionen sozialversicherungspflichtig Beschäftigte den Betrieben des IAB-Betriebspanels zugeordnet werden können. Grundsätzlich sprechen eine Reihe von Vorteilen für die Verknüpfung von Firmendaten und Personendaten. Insbesondere kann bei Betriebsbefragungen auf die Erhebung von Durchschnittsangaben zu Löhnen (bspw. nach Qualifikationsgruppen, Geschlecht, Beruf) verzichtet werden. Durch die Verknüpfung mit personenbezogenen Daten können die in einem Betrieb beschäftigten Arbeitskräfte trotzdem nach diesen Merkmalen differenziert werden. Damit ergeben sich wesentlich differenziertere Analyse-möglichkeiten als mit herkömmlichen Betriebsbefragungen (siehe Abowd and Karmarz 1999).

Eine interessante Vernetzungsmöglichkeit von Befragungsdaten mit der Beschäftigtenstatistik der Bundesanstalt für Arbeit wird in einem Projekt des MPI-Instituts für Bildungsforschung in Berlin erprobt.<sup>26</sup> In diesem Projekt sollen die im Rahmen der Lebensverlaufsstudien (vgl. dazu Abschnitt 2.1.1) erhobenen Befragungsdaten mit den aus der Beschäftigtenstatistik der Bundesanstalt für Arbeit gewonnen Verlaufsdaten über Erwerbsverläufe der entsprechenden Personen verknüpft werden. Ein Ziel dieses Projekts besteht auch darin, durch den Vergleich von retrospektiv erhobenen Befragungsdaten über individuelle Erwerbsverläufe und den im Meldeverfahren generierten Daten der Beschäftigtenstatistik Anhaltspunkte über die Validität von Befragungsdaten zu gewinnen.

Vielversprechende Vernetzungsmöglichkeiten bestehen auch bezüglich der von der Bundesanstalt für Arbeit im Rahmen ihres Verwaltungshandelns erhobenen (prozessproduzierten) Daten. Die „Förderstatistik“ ließe sich über die Sozialversicherungsnummer mit der Beschäftigten- und Leistungsbezieherstatistik verknüpfen. Dadurch könnten informative Ereignisdaten über Erwerbsverläufe und die Entwicklung der Erwerbseinkommen von Teilnehmern an der beruflichen Weiterbildung generiert werden. Damit könnte der gravierende Mangel an Mikrodaten in einem wichtigen Bereich der Arbeitsmarktforschung deutlich reduziert werden. Gerade in diesem Bereich spielt die Vernetzung von Daten eine entscheidende Rolle, da man für wissenschaftliche Evaluationsstudien nicht nur Angaben zu den geförderten Teilnehmern an einer bestimmten Maßnahme, sondern auch Informationen zu einer Kontrollgruppe von Nichtteilnehmern benötigt.

Ähnlich könnte auch für andere Bereiche der Arbeitsförderung ohne zusätzliche kostenintensive Befragungen durch Vernetzung der vorliegenden prozessproduzierten Daten der Arbeitsverwaltung verfahren werden. Ein Beispiel dafür stellt die in einem Pilotprojekt zwischen IAB und ZEW durchgeführte Vernetzung von prozessproduzierten Daten der Bundesanstalt für Arbeit über Überbrückungsgeldempfänger mit den ZEW-Creditreform Unternehmensdaten dar (Pfeiffer und Reize 2000). Ein weiteres Beispiel für die Vernetzung von pro-

---

<sup>26</sup> Vgl. <http://www.mpib-berlin.mpg.de/BAG/projects/project1-1.htm>.

zess-produzierten Daten und Befragungsdaten ist die Studie von Gilberg et al. (1999), die auf der Verknüpfung einer Stichprobe des Bestands der Arbeitslosengeld- und Arbeitslosenhilfeempfänger sowie eine Stichprobe von Abgängern aus Arbeitslosigkeit mit einer eigenen Befragung dieser Arbeitslosen basiert.

Eine weitere Vernetzungsmöglichkeit ist organisatorischer Art. In den einzelnen Landesämtern sind häufig nur die Mikrodaten für das jeweilige Bundesland verfügbar. Die partiell bereits praktizierten Möglichkeiten, in Statistischen Landesämtern wissenschaftliche Analysen mit den Mikrodaten der Amtlichen Statistik durchzuführen, sind in diesen Fällen auf das jeweilige Bundesland beschränkt. Für wissenschaftliche Analysen sind aber häufig die Daten für das gesamte Bundesgebiet von Interesse. Es wäre daher aus logistischen Gründen eine wesentliche Verbesserung, wenn in den einzelnen Landesämtern ein Zugriff auf die bundesweit erhobenen Daten möglich wäre.

## 2.4 Überschneidungen

Aufgrund der bisher fehlenden Vernetzung der einzelnen Erhebungen in den Bereichen Erwerbstätigkeit und Einkommen bestehen in der Amtlichen Statistik zahlreiche Überschneidungen. Diese betreffen den Fragekatalog, den Kreis der Befragten und die Auftraggeber der Erhebungen.

Angaben zur Erwerbstätigkeit und zum Einkommen werden in unterschiedlicher Abgrenzung und Definition in der Amtlichen Statistik an mehreren Stellen erhoben. Die umfassendste Erhebung zur Erwerbstätigkeit erfolgt im Mikrozensus. Andere Erhebungen erfassen meist nur das Produzierende Gewerbe und bestimmte Dienstleistungen. Die umfassendste Erhebung zum Einkommen erfolgt in der Einkommens- und Verbrauchsstichprobe, rudimentär wird das Einkommen aber auch im Mikrozensus erhoben. Eine detaillierte Erfassung des Erwerbseinkommens erfolgt auch in der Gehalts- und Lohnstrukturerhebung, allerdings beschränkt auf das Produzierende Gewerbe und bestimmte Dienstleistungen.

Daten zur sozialversicherungspflichtigen Beschäftigung basieren auf der Beschäftigtenstatistik der Bundesanstalt für Arbeit. Diese erfasst nur einen Teil der gesamten Erwerbstätigkeit, wie sie vom Statistischen Bundesamt auf der Basis des Mikrozensus ausgewiesen wird. Statistiken zur registrierten Arbeitslosigkeit basieren auf den prozess-produzierten Daten der Bundesanstalt für Arbeit. Diese unterscheiden sich von den auf der Basis des Mikrozensus und einem unterschiedlichen Konzept erhobenen Statistiken zur Arbeitslosigkeit. Dieses Konzept entspricht internationalen Konventionen und ist die Basis für internationale Vergleiche, bspw. der sogenannten „standardisierten Arbeitslosenquote“ der OECD.

Überschneidungen bestehen aber auch zwischen der Amtlichen Statistik und anderen Erhebungen. Ein aktuelles Beispiel dafür ist die vom Statistischen Bundesamt durchgeführte Erhebung der ersten drei Wellen der deutschen Stichprobe des Europäischen Haushaltspanels (ECHP), das in der Struktur dem deutschen Sozio-ökonomischen Panel sehr ähnlich ist. Der zusätzliche Nutzen der deutschen Stichprobe des ECHP für die Wissenschaft erscheint eher bescheiden. Da das Statistische Bundesamt diese Erhebung aus rechtlichen Gründen nicht weiterführen kann, werden die weiteren geplanten Wellen auf dem SOEP basieren. Damit ist die Kontinuität bezüglich der Stichprobe und dem Design der Erhebung – ein wichtiges Kriterium bei Panelanalysen – nicht gegeben.

Eine andere Form der Überschneidung stellt die unverbundene Datenerhebung im Auftrag der Ministerien, der Bundesanstalt für Arbeit und anderer öffentlicher Stellen dar. Ein Beispiel dafür sind Erhebungen zur geringfügigen Beschäftigung. Diese wird im Mikrozensus, aber auch im SOEP regelmäßig erfasst. Da nach Meinung des Bundesministeriums für Arbeit und Sozialordnung (und der meisten Wirtschaftsforscher) die geringfügige Beschäftigung durch den Mikrozensus nur zu einem geringen Teil erfasst wird, wurde durch ein privates Forschungsinstitut eine eigene Befragung dazu durchgeführt, die auch stark abweichende Ergebnisse erbrachte. Dies hat jedoch nicht dazu geführt, dass die entsprechende Frage im Mikrozensus angepasst wurde.

Häufig werden auch durch Befragungen Informationen erhoben, die zum Teil bereits in Form prozess-produzierter Daten vorliegen. Dies betrifft beispielsweise Verlaufserhebungen zur Arbeitslosigkeit, bei denen retrospektiv Angaben über die zeitliche Lage individueller Arbeitslosigkeitsspannen oder die Teilnahme an durch die Bundesanstalt finanzierte Weiterbildungsmaßnahmen erhoben werden. Derartige Angaben könnten wesentlich genauer aus den prozess-produzierten Daten der Bundesanstalt übernommen werden. Darüber hinaus ließe sich die Anzahl der zu erhebenden Fragen reduzieren.

Generell ist zu vermuten, dass im Rahmen von Forschungsprojekten im Auftrag von Ministerien, der Bundesanstalt für Arbeit und anderen öffentlichen Auftraggebern Daten zu bestimmten Fragestellungen mehrfach erhoben werden. Diese Mehrfacherhebungen könnten dadurch reduziert werden, dass ein mit der Erhebung von Daten verbundener Forschungsauftrag mit der Verpflichtung verknüpft wird, die Daten nach Abschluss des Forschungsprojekts für wissenschaftliche Zwecke in einer dafür geeigneten Form zur Verfügung zu stellen. Dies geschieht bereits bei Projekten, die durch die Deutsche Forschungsgemeinschaft finanziert werden. Die Weitergabe der in geeigneter Form aufbereiteten (bspw. faktisch anonymisierten) Daten an die Wissenschaft könnte durch eine oder mehrere Clearing-Stellen (bspw. dem Zentralarchiv, ZUMA oder auch private Unternehmen) erfolgen. Die Finanzierung der Grundkosten für die mit der Weitergabe verbundenen Datenaufbereitung sollte vom öffentlichen Auftraggeber des Forschungsprojekts, die Zusatzkosten von den Datennutzern getragen werden.

### **3 Auftraggeber für die Datenproduktion**

In Tabelle 2 sind die Auftraggeber für die oben näher beschriebenen Datensätze aufgelistet. Im Bereich der Amtlichen Statistik ist der Bund Auftraggeber. Auf dem Gebiet von Ausbildung und dualem Berufsausbildungssystem ist das BIBB und damit das vorgeschaltete Ministerium (BMBF) der Auftraggeber. Das BMBF ist auch Auftraggeber für die regelmäßigen HIS-Erhebungen über Hochschul- und Hochschulabsolventen. Die Bundesanstalt für Arbeit und das ihr zugeordnete IAB sind Auftraggeber für eine Reihe von regelmäßigen Erhebungen, insbesondere für das IAB-Betriebspanel, sowie eine Vielzahl von Spezialerhebungen im Bereich Erwerbstätigkeit. Auftraggeber des SOEP ist die Wissenschaft selbst. Auftraggeber für Befragungen zu bestimmten Themenkomplexen im Bereich Erwerbstätigkeit und Einkommen sind manchmal auch einzelne Wissenschaftler und wissenschaftliche Institute im Rahmen von Forschungsprojekten, die durch Wissenschaftsorganisationen oder Stiftungen finanziert werden. Private Auftraggeber spielen bei der Produktion von Daten zu Erwerbstätigkeit und Einkommen, die auch von der Wissenschaft genutzt werden können, bisher keine Rolle.

Tabelle 2: Überblick über die Auftraggeber

Datensatz	Auftraggeber
ALLBUS	Universitäten
Amtliche Statistik: Mikrozensus, EVS, Zeitbudgeterhebung, GLS	Bund
Betriebsdatei, Beschäftigtenstatistik	Bundesanstalt für Arbeit
BIBB-IAB Erhebung	BIBB/BMBF
CoARB	Bundesanstalt für Arbeit
ECHP	Eurostat
HIS-Daten	HIS/BMBF
IAB-Beschäftigtenstichprobe	IAB
IAB-Betriebspanel	IAB
Lebensverlaufsstudie	Max-Planck-Institut für Bildungsforschung
SOEP	Wissenschaft (SOEP Gruppe)
Sozialhilfestatistik	Bund

Anmerkung: Die Datensätze sind in alphabetischer Reihenfolge sortiert.

## 4 Finanzierung

Daten auf dem Gebiet der Erwerbstätigkeit und Einkommen werden fast ausschließlich durch öffentliche Gelder finanziert. Das SOEP wird seit 1990 als eigenes DFG-Projekt gefördert. Auf Beschluß der Bund-Länder Kommission für Bildungsplanung und Forschungsförderung (BLK) ist die Finanzierung bis mindestens einschließlich des Jahres 2002 gesichert. Die Finanzmittel werden auf Basis von Begutachtungen durch die DFG als Sonderfinanzierung von der BLK zur Verfügung gestellt. Aufgebracht werden die Mittel jeweils zur Hälfte von Bund und Ländern, wobei das Sitzland Berlin 25% des Länderanteils trägt.<sup>27</sup> Die Kosten der Erhebung von Bundesstatistiken (Mikrozensus, EVS) werden generell vom Bund getragen. Die Erhebung von HIS und BIBB werden ausschliesslich mit öffentlichen Geldern finanziert. Die Kosten der Erhebung des IAB Betriebspanel belaufen sich auf ca. 2 Millionen DM pro Jahr und werden vom IAB und dem Europäischen Sozialfond getragen. Die Beschäftigtenstichprobe wird vom IAB, DFG und dem Europäischen Sozialfond finanziert. Bundesministerien sind ebenfalls Auftraggeber einer Reihe von Erhebungen. Beispielsweise wird das Mannheimer Innovationspanel vom BMBF finanziert.

Universitäten verfügen im allgemeinen über keine Finanzmittel für die Erhebung von großen Einzeldatensätzen. Selbstfinanzierte Datensätze beschränken sich auf sehr kleine Stichproben. Beispielsweise enthält der Firmendatensatz von Kraft (1991) nur 60 Firmen. Die

<sup>27</sup> ([http://www.diw.de/soep/info\\_flyer\\_de.html#orga](http://www.diw.de/soep/info_flyer_de.html#orga))

Erhebung von Hart und Shipmann (1991) enthält 20 Firmen. Wissenschaftsstiftungen finanzieren nur in Ausnahmefällen Datenerhebungen. Eine der wenigen Ausnahmen ist das Hannoveraner Firmenpanel, das von der Volkswagenstiftung finanziert wurde. Allgemein sollte die Bereitstellung von Daten für die Wissenschaft durch zentral finanzierte Infrastruktureinrichtungen finanziert werden, wenn die Bereitstellung derartiger Daten nicht bereits Grundaufgabe der Amtlichen Statistik ist (vgl. Müller 1999, S. 24 f.).

Dieser Grundsatz wurde bereits ansatzweise bei den Haushaltsdaten der Amtlichen Statistik realisiert. Wie bereits erwähnt, wurden die faktisch anonymisierten Einzeldaten des Mikrozensus, der Einkommens- und Verbrauchsstichprobe, der deutschen Stichprobe der ersten drei Wellen des Europäischen Haushaltspanels und der Zeitbudgeterhebung 1991/92 im Rahmen eines Pilotprojekts zwischen dem Statistischen Bundesamt und dem BMBF zu Grenzkosten an die Wissenschaft für Forschungszwecke weitergegeben. Im Rahmen eines Übereinkommens zwischen dem Statistischen Bundesamt und der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS) wurden weitere Erhebungen des Mikrozensus in faktisch anonymisierter Form der Wissenschaft zu den gleichen Bedingungen zur Verfügung gestellt. Leider konnte bisher keine entsprechende Finanzierung für die Einkommens- und Verbrauchsstichprobe 1998 gefunden werden. Die faktisch anonymisierten Daten der EVS 98 werden zur Zeit vom Statistischen Bundesamt zu Preisen zwischen DM 1.620 für 100 Eingabefelder und DM 7.130 für den gesamten Fragekatalog angeboten. Bei Bedarf an einer größeren Anzahl von Variablen dürften diese Preise für viele Wissenschaftler, insbesondere an den Universitäten, prohibitiv wirken. Die Finanzierung dieser für die Wissenschaft wichtigen Datenbasis sollte daher hohe Priorität haben.

Tabelle 3: Überblick über die Finanzierung

Datensatz	Finanzierung
ALLBUS	GESIS/BMBF
Amtliche Statistik: Mikrozensus, EVS, Zeitbudgeterhebung, GLS	Bund
Betriebsdatei, Beschäftigtenstatistik	Bundesanstalt für Arbeit
BIBB-IAB Erhebung	BIBB/BMBF (=Bund)
CoARB	Bundesanstalt für Arbeit (=Bund)
ECHP	Eurostat/Bund
HIS-Daten	HIS/BMBF (=Bund)
IAB-Beschäftigtenstichprobe	IAB, DFG, Europäischer Sozialfond
IAB-Betriebspanel	Bundesanstalt für Arbeit, Europäischer Sozialfond
Lebensverlaufsstudie	Wissenschaft
SOEP	DFG, BLK
Sozialhilfestatistik	Bund

Anmerkung: Die Datensätze sind in alphabetischer Reihenfolge sortiert.

## 5 Rechtsgrundlagen

Es existiert ein individuelles Grundrecht auf Datenschutz, durch das individuelle Daten vor der Weitergabe an Dritte geschützt werden. Die Rechtsgrundlagen der Amtlichen Statistik im Bereich Erwerbstätigkeit und Einkommen sind im Bundesstatistikgesetz (BDSG) geregelt (siehe Gnos 1999). Relevante datenschutzrechtliche Bestimmungen finden sich auch im Sozialgesetzbuch (SGB, Buch X). Die Weitergabe von Sozialdaten zu Forschungszwecken an nicht dem Sozialgeheimnis unterworfenen Stellen („Datenübermittlung“) bedarf nach §75 SGB X der Genehmigung. Bei prozess-produzierten Daten der verschiedenen Sozialversicherungsbereiche gibt es verschiedene Sonderbestimmungen, bspw. bei der Krankenversicherung. Die Datenschutzaufsicht liegt beim Bundesbeauftragten bzw. den Landesbeauftragten für den Datenschutz, je nachdem ob es sich um Bundes- oder Landesdaten handelt. Für privat erhobene Daten gelten die erwähnten rechtlichen Bestimmungen nicht. Allerdings besteht zwischen Erhebungsinstituten ein informelles Übereinkommen, dass auch im Rahmen der Auftragsforschung erhobene Daten Datenschutzbestimmungen berücksichtigt werden.

## 6 Datenschutzregelungen

Daten der Amtlichen Statistik (einschließlich den sogenannten Sozialdaten) dürfen nur bei der Erfüllung strenger datenschutzrechtlicher Bestimmungen an Dritte weitergegeben werden. Im allgemeinen muss eine Identifikation der betreffenden Untersuchungseinheiten aus den Daten ausgeschlossen sein. Bezüglich der Verwendung dieser Daten für wissenschaftliche Zwecke bestehen seit 1987 aber Sonderbestimmungen („Wissenschaftsprivileg“). Die Weitergabe von Einzeldaten an die Wissenschaft ist in §16/6 BDSG geregelt. Danach dürfen für Forschungszwecke vom Statistischen Bundesamt und den statistischen Ämtern der Länder personenbezogene Daten an Hochschulen oder sonstige unabhängige Forschungseinrichtungen übermittelt werden, wenn diese Daten „faktisch anonymisiert“ sind. Dies bedeutet, dass diese Daten nur mit einem unverhältnismäßig großen Aufwand an Zeit und Kosten den betroffenen Personen zugeordnet werden können (Gnos 1999). Voraussetzung für die Übermittlung der Daten ist der Abschluss eines Vertrages. Die Verwendung dieser Daten ist zweckgebunden. In der Regel setzt dies ein konkretes Forschungsprojekt voraus, das auch den Umfang der weitergegebenen Daten (bezüglich der Variablenauswahl) festlegt. Nach Abschluss des wissenschaftlichen Vorhabens müssen die übermittelten Daten gelöscht werden.

Wie oben erwähnt, werden im Bereich der Personen- und Haushaltsdaten der Amtlichen Statistik bereits faktisch anonymisierte Daten des Mikrozensus, der Einkommens- und Verbrauchsstichprobe, der deutschen Stichprobe des Europäischen Haushaltspanels und der Zeitbudgeterhebung durch das Statistische Bundesamt für Forschungszwecke bereitgestellt. Auch faktisch anonymisierte Daten der Beschäftigungs- und Leistungsstatistik („Sozialdaten“) konnten der Wissenschaft durch zwei Stichproben der I<sub>A</sub>BS bereits zur Verfügung gestellt werden.

Schwieriger ist die faktische Anonymisierung von Betriebs- und Unternehmensdaten, da die Identifikation der Untersuchungseinheiten aus den Daten leichter möglich und wahr-

scheinlicher ist. Als Begründung für diese Einschätzung wird in der Literatur angeführt, dass die Möglichkeit der Re-identifikation bei Betriebs- und Unternehmensdaten über leicht verfügbare externe Zusatzinformationen größer (d.h. die Kosten geringer) und die Anreize zur Re-identifikation größer als bei personenbezogenen Mikrodaten wären (Gnoss 1999). Eine wirksame Anonymisierung wäre daher im Bereich von Wirtschaftsdaten häufig mit einem erheblichen Informationsverlust verbunden (Neubauer 1999, S. 105), so dass das Analysepotenzial stark eingeschränkt wird. Beispielsweise werden externen Wissenschaftlern durch das ZEW faktisch anonymisierte Datensätze des Mannheimer Innovationspanels (MIP) für Forschungszwecke zur Verfügung gestellt. Dabei wird eine Reihe von quantitativen Variablen (Beschäftigung, Umsatz) verschlüsselt, so dass diese Variablen nur qualitativ analysiert werden können. Für das IAB-Betriebspanel stehen bisher offiziell externen Wissenschaftlern keine faktisch anonymisierten Daten zur Verfügung. Die Möglichkeit der Anonymisierung wird noch geprüft.

Aus diesem Grund sind neben der faktischen Anonymisierung von Wirtschaftsdaten alternative Möglichkeiten der Nutzung von Mikrodaten durch die Wissenschaft zu prüfen. Eine Möglichkeit besteht darin, dass der Wissenschaftler in der Institution, die über die Einzeldaten verfügt, an einem speziell gesicherten Computer die wissenschaftliche Auswertung durchführt. Bei Daten, die nur für Analysen im Auftrag der über die Daten verfügenden Institution verwendet werden dürfen, könnte für die Projektdauer ein formales Beschäftigungsverhältnis zwischen der aufnehmenden Institution und dem Wissenschaftler geschlossen werden. Dies wird beispielsweise in den USA als „One-Dollar-Contract“ in bestimmten Fällen praktiziert. Die Nutzung derartiger Regelungen könnte erleichtert werden, wenn über regionale Rechenzentren eine dezentrale Nutzung dieser Daten ermöglicht würde, wie dies beispielsweise auch bei den Wirtschaftsdaten des Bureau of the Census in den USA zunehmend praktiziert wird. Eine weiterführende Möglichkeit des Zugangs zu nicht anonymisierten Mikrodaten stellt das sogenannte „Fernrechnen“ dar, bei dem dazu berechtigte Wissenschaftler über das Internet direkt über einen dafür speziell gesicherten Rechner auf die Mikrodaten zugreifen können. Dabei kann durch hardware- und software-technische Beschränkungen mit relativ einfachen Mitteln ausgeschlossen werden, dass eine Identifikation von Einzeldaten erfolgen kann. Diese Möglichkeit der Sicherung der datenschutzrechtlichen Bestimmungen wird bereits bei Personendaten genutzt, beispielsweise im Rahmen der „Luxembourg Income Study“.

## **7 Informationsverbreitung**

In den letzten Jahren hat die starke Verbesserung der Computerleistung und die rasante Entwicklung des Internet zu einer zunehmenden Informationsverbreitung geführt (siehe Nowack 1997). Eine Vielzahl von aggregierten Daten kann mittlerweile über das Internet heruntergeladen werden. Im Gegensatz beispielsweise zu den USA können Mikrodaten in Deutschland bisher noch nicht über das Internet heruntergeladen werden. Die Verfügbarkeit von Mikrodatensätzen für Übungszwecke in der Lehre wäre aber wünschenswert.

Trotz der zunehmenden Verfügbarkeit von aggregierten Daten im Internet sind Printmedien immer noch Schwerpunkt der Informationsverbreitung der Amtlichen Statistik. Häufig nachgefragte zeitnahe Daten, beispielsweise zu Beschäftigten nach Qualifikation und Branche in tiefstmöglicher Gliederung auf Basis des Mikrozensus oder der Beschäftigtenstatistik, sind

weder in Printmedien noch im Internet verfügbar. Sonderauswertungen können beim Statistischen Bundesamt oder bei der Bundesanstalt für Arbeit direkt in Auftrag gegeben werden (siehe Gnos 1999), sind aber nicht ganz billig und außerdem benötigt deren Erledigung häufig relativ viel Zeit.

## 7.1 Informationsverbreitung über Printmedien

Tabellenauswertungen amtlicher Statistiken zu Erwerbstätigkeit und Einkommen werden in den Fachserien 1, 15 und 16 publiziert. Von zentraler Bedeutung sind die Fachserie 1, Reihe 4 (Erwerbstätigkeit) und Fachserie 15 (Einkommen).

Die Ergebnisse des Mikrozensus sind überwiegend nur in Printmedien verfügbar, werden aber relativ zeitnah veröffentlicht. Im Einzelnen sind die Ergebnisse des Mikrozensus in der Fachserie 1, Reihe 4.1.1: „*Stand und Entwicklung der Erwerbstätigkeit*“ und Reihe 4.1.2: „*Bevölkerung und Erwerbstätigkeit*“ zusammengefasst. Hier werden überwiegend zweidimensionale Kreuztabellen ausgewiesen. Wünschenswert ist der weitere Ausbau aggregierter Tabellenauswertungen auf Basis des Mikrozensus.

Im Bereich der Lohn- und Gehaltsstatistiken veröffentlicht das Statistische Bundesamt alle fünf Jahre große und sehr detaillierte Tabellenbände. Tabellenauswertungen zu Struktur und Entgelt der sozialversicherungspflichtig Beschäftigten sind in Reihe 1, Fachserie 4.2.1 und 4.2.2 ausgewiesen. Die sozialversicherungspflichtig Beschäftigten sind nach Branche, Geschlecht, Vollzeit/Teilzeit sowie Angestellte/Arbeiter untergliedert. Kreuzklassifikationen nach der beruflichen Qualifikation fehlen jedoch. Für detailliertere Analysen sind Wissenschaftler deswegen auf Sonderauswertungen der Beschäftigtenstatistik angewiesen.

Detaillierte tabellarische Auswertungen zum Einkommen auf Basis der Einkommens- und Verbrauchsstichprobe werden vom Statistischen Bundesamt in der Fachserie 15: Wirtschaftsrechnungen publiziert. Insbesondere wird hier alle fünf Jahre die Entstehung und Verwendung des Einkommens nach verschiedenen Haushaltstypen nachgewiesen. In Tabellenform werden in dieser Fachserie auch die Ergebnisse der laufenden Wirtschaftsrechnungen in Vierteljahresberichten und wesentlich detaillierter in den Jahresberichten publiziert.

Die Informationsverbreitung in Printmedien auf Basis der Amtlichen Industriestatistik (Monatsberichte, Kostenstrukturerhebung) und der VGR ist äußerst umfangreich. Für die amtliche Industriestatistik würden vierteljährliche Publikationen genügen, da die monatliche Variabilität der Beschäftigung generell sehr gering ist. Diese werden jedoch für wissenschaftliche Zwecke selten eingesetzt, da wichtige Bereiche wie unternehmensbezogene Dienstleistungen fehlen.

## 7.2 Informationsverbreitung über das Internet

Das mit Abstand bedeutendste Internetangebot wird vom Statistischen Bundesamt bereitgestellt. Seit 1996 bietet der Zeitreihenservice des Statistischen Bundesamtes einen Onlinezugang zur Datenbasis des Statistischen Informationssystem des Bundes (STATIS-BUND) an.<sup>28</sup> Der Zugang ist kostenpflichtig. Das Angebot umfaßt lange Jahresreihen, die zum Teil bis

<sup>28</sup> Siehe hierzu <http://www-zr.statistik-bund.de/zeitreih/home.htm>

1950 zurückreichen und kurze Monats- und Vierteljahresreihen.<sup>29</sup> Auf dem Gebiet der Erwerbstätigkeit sind ca. 1000 Zeitreihen verfügbar. Im Einzelnen werden die Anzahl der Erwerbstätigen, Erwerbspersonen und der Beschäftigten nach verschiedenen Merkmalen (ein-stellige Wirtschaftsklassen, Alter, Geschlecht, berufliche Qualifikation, Berufsgruppen) untergliedert. Datenbasis sind die Beschäftigtenstatistik und der Mikrozensus. Das Internetangebot von Daten zur Erwerbstätigkeit und zu Entgelten sowie Arbeitskosten auf Basis der VGR, laufenden Verdiensterhebungen, Kostenstrukturerhebungen und der amtlichen Industriestatistik ist zufriedenstellend. Eher dürftig ist dagegen das Angebot von Zeitreihen auf Basis des Mikrozensus im Internet. Lange Reihen, wie beispielsweise die Anzahl der Selbstständigen nach Wirtschaftszweigen sind nur den Printmedien (siehe Reihe 1 Fachserien 4.1.1 und 4.1.2) zu entnehmen, nicht aber online erhältlich. Ähnlich dürftig ist das Internetangebot des Statistischen Bundesamtes bezüglich Statistiken zum Einkommen auf der Basis der Wirtschaftsrechnungen.

Wünschenswert wäre die kostenlose Bereitstellung eines Katalogs wichtiger Statistiken zur Erwerbstätigkeit und zum Einkommen im Internet. Beispielsweise bietet das U.S Bureau of Labor Statistics einen kostenlosen Zugang zu mehreren zehntausend Daten zu Erwerbstätigkeit und Einkommen an.<sup>30</sup>

Für Bundesstatistiken, die dezentral bei den Statistischen Landesämtern vorliegen (Amtliche Industriestatistik, Verwaltungsdaten), ist das Internetangebot der Statistischen Landesämter interessant. Führend ist das Angebot des Statistischen Landesamts Baden-Württemberg (LIS).<sup>31</sup> Für interessierte Institutionen, Firmen, Verbände usw. besteht auch die Möglichkeit, sich direkt über Telnet an das LIS anzuschließen. Das Angebot ist allerdings von der Wissenschaft bisher kaum genutzt worden.

In jüngster Zeit hat auch die Bundesanstalt für Arbeit ihr Internetdatenangebot erheblich ausgebaut.<sup>32</sup> Ausgewählte Tabellen aus den Amtlichen Nachrichten der BA und der Heftreihe "Arbeitsmarkt in Zahlen" werden kostenlos für das letzte Jahr zur Verfügung gestellt, teilweise sind Zeitreihen für den Zeitraum 1991- 2000 verfügbar. Zeitreihen über die Anzahl von Arbeitssuchenden, Arbeitslosen, Arbeitslosenquoten, offenen Stellen, Zugängen an Arbeitssuchenden, Arbeitsvermittlungen, Leistungsempfängern, Kurzarbeitern und Arbeitsbeschaffungsmaßnahmen sind verfügbar. Im Vergleich zum Bureau of Labor Statistics (siehe oben) ist das Datenangebot aber noch ausbaufähig.

## 8 Verfügbarkeit von Einzeldaten

Tabelle 4 enthält eine Zusammenfassung der oben beschriebenen Mikrodatensätze mit Angaben darüber, für welche Perioden diese Datensätze verfügbar sind und ob sie an die Wissenschaft weitergegeben werden. Die Verfügbarkeit bezieht sich nur auf die Nutzung für wissenschaftliche Zwecke durch Wissenschaftler an Universitäten und sonstigen wissenschaftlichen

---

<sup>29</sup> Siehe zu der Beschreibung hierzu ausführlich: <http://www-zr.statistik-bund.de/zeitreih/hilfe/info.htm>.

<sup>30</sup> Siehe hierzu <http://www.bls.gov>

<sup>31</sup> Siehe hierzu: <http://www.statistik.baden-wuerttemberg.de/SRDB/>

<sup>32</sup> Siehe hierzu <http://www.arbeitsamt.de/hst/services/archiv/aktuell/download.html>

Einrichtungen. Darüber hinaus kann es noch weitere Einschränkungen geben, wie beispielsweise die Beschränkung der Weitergabe an Inländer (vgl. dazu Abschnitt 9.1).

Tabelle 4: Überblick über die Verfügbarkeit und Weitergabe von Mikrodaten

Datensatz	Verfügbarkeit	Weitergabe
ALLBUS	80-98, zweijährig,	auf Antrag beim ZA
Beschäftigtenstatistik	1975 – 1999	keine
BIBB-IAB Erhebung	78/79, 85/85, 91/92, 98/99 wahrscheinlich ab 2002/2003	auf Antrag beim ZA
CoARB	seit 70er Jahre	beschränkt, auf Zustimmung der Bundesanstalt für Arbeit
ECHP	1994 – 1996	faktisch anonymisiert auf Antrag beim Statisti- schen Bundesamt
EVS	1993, 1998	faktisch anonymisiert für 1993 und 1998
GLS	für 1990 und 1995 bei den Landesämtern	Keine Weitergabe, 1 Pilot- projekt mit der niedersächsi- schen Stichprobe
HIS-Daten	90er Jahre	keine Weitergabe
IAB-Beschäftigtenstich- probe	1975 – 1990 1975 – 1995	auf Antrag beim ZA
IAB-Betriebspanel	ab 1993 jährlich	keine Weitergabe, Schalter- stelle
Lebensverlaufsstudie	verschiedene Erhebungen 1981-83, 1985/86 , 1987/88, 1989	auf Antrag beim ZA
Mikrozensus	1989, 91, 93, 95, 96 und 97	faktisch anonymisiert auf Antrag beim Statisti- schen Bundesamt
SOEP	seit 1984 für West-, seit 1990 für Ostdeutschland, jährlich	auf Antrag beim DIW
Sozialhilfestatistik	1994 – 1997	Weitergabe der faktisch anonymisierten Stichprobe geplant

*Anmerkungen:* Die Datensätze sind in alphabetischer Reihenfolge sortiert. Bei den Datensätzen der Amtlichen Statistik (Mikrozensus, EVS, Sozialhilfestatistik) bezieht sich die Verfügbarkeit auf die

faktisch anonymisierten Datensätze, die an die Wissenschaft weitergegeben werden, nicht auf die verfügbaren Erhebungen insgesamt. ZA = Zentralarchiv für empirische Sozialforschung, Köln.

## 9 Weitergabe von Einzeldaten

### 9.1 National/international

Der Zugang zu Personen- und Haushaltsdaten der Amtlichen Statistik ist durch ein Pilotprojekt zwischen dem Statistischen Bundesamt und dem Bundesministerium für Bildung und Forschung (BMBF) zur Bereitstellung von faktisch anonymisierten Daten an die Wissenschaft erheblich erleichtert worden (vgl. hierzu auch Schimpl-Neimanns 2000, Köhler, et al. 2000). Im Rahmen dieses Pilotprojekts hat das BMBF die Grundkosten für die Bereitstellung folgender faktisch anonymisierter Mikrodaten:

- Mikrozensus 1995, 1996 und Mikrozensus 1997
- Einkommens- und Verbrauchsstichprobe 1993
- Deutsche Stichprobe des Europäischen Haushaltspanels 1994-96
- Zeitbudgeterhebung 1991/92

übernommen. Diese sogenannten „Scientific-Use Files“ können im Rahmen eines Forschungsprojekts beim Statistischen Bundesamt bezogen werden.<sup>33</sup> Dabei ist zu beachten, dass Scientific Use Files nach dem Inländerkonzept nur von unabhängigen inländischen Forschungseinrichtungen bestellt werden können. Die Mikrozensus-Richtlinien beschränken die Nutzung der Daten auf Deutschland. Ausländische Forscher haben jedoch die Möglichkeit, bei ZUMA mit den Daten nach besonderer Verpflichtung zu arbeiten.<sup>34</sup>

Da das Pilotprojekt mit dem BMBF zeitlich befristet war und die Wissenschaft insbesondere am Mikrozensus starkes Interesse gezeigt hat, haben die Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen (GESIS) und das Statistische Bundesamt eine Ausweitung der für die Forschung zugänglichen Datenbasis des Mikrozensus um die Scientific-Use-Files der Erhebungen 1989, 1991, 1993 und 1997 vereinbart. Die GESIS übernimmt die Grundkosten für die Bereitstellung der Mikrodatenfiles, die ebenfalls für eine Bearbeitungsgebühr von jeweils 130 DM beim Statistischen Bundesamt bestellt werden können. Die Lieferung erfolgt auf einem maschinenlesbaren Datenträger (u.a. CD-ROM) in der Regel einen Monat nach Abschluß des Vertrages. Die Betreuung und die Information der Nutzer werden vom Statistischen Bundesamt in Kooperation mit dem Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim durchgeführt. Ausführliche Informationsunterlagen über vorhandene Variablen und verwendete Systematiken können vom Statistischen Bundesamt bezogen werden. Von ZUMA wird zudem ein Set Up zum Einlesen in SPSS bereitgestellt.

Ergänzend zu Scientific Use Files wird zur Zeit in einzelnen Bereichen der Amtlichen Statistik geprüft, ob auch sogenannte Public Use Files angeboten werden können. Hierunter versteht man Standardprodukte mit absolut anonymisierten Mikrodaten nach § 16 Abs. 1 Nr.

<sup>33</sup> Siehe hierzu: [http://www.statistik-bund.de/micro/micro\\_c2.htm](http://www.statistik-bund.de/micro/micro_c2.htm)

<sup>34</sup> Siehe hierzu Zuma:  
[http://www.zuma-mannheim.de/data/microdata/scientific\\_use\\_file/scientific\\_use\\_files.htm](http://www.zuma-mannheim.de/data/microdata/scientific_use_file/scientific_use_files.htm)

4 Bundesstatistikgesetz, die nicht nur die Wissenschaft, sondern alle Interessierten beziehen können. In Vorbereitung sind Public Use Files der Zeitbudgeterhebung 1991/92 und der 25%-Stichprobe der Sozialhilfestatistik.

Die anderen in der Tabelle 4 aufgelisteten personenbezogenen Datensätze, die an die Wissenschaft weitergegeben werden, können direkt bei dem DIW in Berlin (SOEP) oder beim Zentralarchiv (IAB-Beschäftigtenstichprobe und BIBB/IAB Erhebung) bestellt werden. Das Verfahren ist ähnlich bei den Daten der Amtlichen Statistik. Allerdings wird eine Teilstichprobe (von fast 100%) des SOEP auch an Wissenschaftler mit ausländischem Wohnsitz bzw. Arbeitsort weitergegeben.

Die Weitergabe von faktisch anonymisierten Betriebs- und Unternehmensdaten wird aus den erwähnten Gründen als problematischer angesehen als die von personenbezogenen Daten (vgl. auch Müller 1999). Wie erwähnt, wird diese Möglichkeit aber trotzdem praktiziert, beispielsweise beim Mannheimer Innovations Panel oder beim NIFA-Panel für den Maschinenbau (vgl. Anhang). Das IAB-Betriebspanel als wichtigste Erhebung betriebsbezogener Mikrodaten auf dem Gebiet der Erwerbstätigkeit steht zur Zeit Wissenschaftlern außerhalb des IAB aus datenschutzrechtlichen Gründen offiziell nicht zur Verfügung. Um aber trotzdem auch externen Wissenschaftlern die Arbeit mit dem IAB-Betriebspanel zu ermöglichen, ist eine sogenannte Schalterstelle beim IAB eingerichtet worden. Diese Schalterstelle, über die Berechnungen mit den Mikrodaten des IAB-Betriebspanels auf der Basis von durch externe Wissenschaftler geschriebenen Auswertungsprogrammen durchgeführt werden sollen, wird vom BMBF finanziert. Nach Angabe des zuständigen Projektleiters beim IAB beträgt die Anzahl der Nutzer dieser Schalterstelle zur Zeit ca. 50. Zur Erleichterung der Erstellung von Auswertungsprogrammen wird Interessenten seit Ende 1999 ein Testdatensatz des IAB zur Verfügung gestellt. Auf Basis dieser Testdatensätze soll es möglich sein, eigene Programme zur Auswertung zu erstellen. Der Testdatensatz enthält für jede Welle Variablenabkürzungen und Angaben für 300 Betriebe, jedoch keine richtigen Daten. Die von den einzelnen Wissenschaftlern erstellten Auswertungsprogramme werden dann im IAB mit den Originaldaten durchgeführt. Nach der obligatorischen datenschutzrechtlichen Überprüfung der Schätzergebnisse erfolgt die Übermittlung der Resultate.

Gegen die Schalterstelle sprechen vor allem zwei Gründe: Zum einen erscheint dieses Verfahren nicht operational, da zur Erklärung komplexer empirischer Phänomene in der Regel unterschiedliche ökonomische Spezifikationen getestet werden müssen. Es ist sehr fraglich, ob dazu auch ein wesentlich verbesserter Testdatensatz ausreichen würde. Außerdem ist dieses Verfahren für Wissenschaftler, die ihre Programme potenziellen Konkurrenten offenlegen sollen, bevor sie die Möglichkeit zur Publikation ihrer Forschungsergebnisse hatten, offensichtlich wenig attraktiv.

Betriebs- und unternehmensbezogene Daten der Amtlichen Statistik werden bisher grundsätzlich nicht weitergegeben. Allerdings ist es seit kurzem möglich, im Rahmen von Kooperationsprojekten mit dem Statistischen Bundesamt oder den Statistischen Ämtern der Länder, diese Daten „vor Ort“ für wissenschaftliche Zwecke zu nutzen. Anfang der 90er Jahre wurden in einem Kooperationsprojekt mit dem Niedersächsischen Landesamt für Statistik Mikrodaten aus Erhebungen der Amtlichen Statistik bei Industriebetrieben über die Zeit zu einem Paneldatenschutz verknüpft und - unter strikter Beachtung der datenschutzrechtlichen Vorschriften - für wissenschaftliche Analysen zur Verfügung gestellt (siehe Wagner 1999).

Dieses Projekt hatte Modellcharakter für andere Bundesländer und hat auch international Beachtung gefunden (siehe Wagner 1999).

Insgesamt sollte stärker berücksichtigt werden, dass es kein allgemeingültiges „Modell“ der Datenweitergabe geben kann und sowohl die Datenproduzenten als auch die Wissenschaft verschiedene Möglichkeiten in der Praxis erproben sollten, wie dies sowohl von wissenschaftlicher als auch politischer Seite betont wird (siehe Müller 1999 und Ziegler 1999).

## **9.2 Preisgestaltung**

Sofern die in Tabelle 4 aufgelisteten Daten überhaupt weitergegeben werden, werden nur die Grenzkosten der Weitergabe berechnet. Eine Ausnahme stellt die Einkommens- und Verbrauchsstichprobe 1998 dar. Zur Zeit werden die faktisch anonymisierten Einzeldaten der EVS 1998 noch zu deutlich höheren Preisen (je nach Umfang zwischen 1.620 und 7.130 DM) angeboten. Die Bezugskosten der im Rahmen des Pilotprojekts zwischen dem Statistischen Bundesamt und dem BMBF zur Verfügung gestellten faktisch anonymisierten Daten der Amtlichen Statistik betragen 130 DM je Erhebung. Für die Mikrozensusdaten der Erhebungen 1989, 1991 und 1993 sowie ab Sommer 2000 der Daten des Mikrozensus 1997 gelten die gleichen Sonderkonditionen. Ähnlich gering sind die Kosten des Bezugs des SOEP und der beim Zentralarchiv verfügbaren Mikrodaten, die ebenfalls zu Grenzkosten abgegeben werden.

## **9.3 Wichtige Nutzergruppen**

Mikrodaten werden überwiegend von Wissenschaftlern genutzt. Bei den meisten der in Tabelle 4 aufgelisteten Datensätzen ist dies ex definitione der Fall, da deren Nutzung auf Personen beschränkt ist, die an Universitäten oder sonstigen unabhängigen Einrichtungen beschäftigt sind. Tabellenauswertungen der Amtlichen Statistik und der Bundesanstalt für Arbeit werden von Wissenschaftlern relativ wenig genutzt. Teilweise werden sie im Rahmen der Auftragsforschung von wissenschaftlichen Forschungsinstituten für deskriptive Darstellungen verwendet. Aggregierte Daten zu Erwerbstätigkeit und Einkommen werden darüber hinaus auch von privaten Forschungsinstituten und –abteilungen von Banken und größeren Unternehmen für Konjunkturprognosen genutzt. Diese Daten werden auch online von privaten Unternehmen (z.B. Datastream) angeboten.

Der Nutzerkreis für die in Tabelle 4 zusammengestellten Mikrodaten ist ziemlich heterogen. Das SOEP als die wichtigste Datenbasis für die Wissenschaft wird wegen des sehr umfangreichen Fragenkatalogs von empirisch orientierten Wirtschafts- und Sozialforschern mit unterschiedlichen Spezialdisziplinen genutzt. Zum Teil gilt dies auch für den Mikrozensus, während die I<sub>A</sub>BS aufgrund der geringen Anzahl der erfassten Merkmale bisher eher für ökonomische Analysen genutzt wurde. Dies zeigt sich auch bei der Struktur der Teilnehmer an sogenannten Nutzerkonferenzen. Diese finden beispielsweise für SOEP Nutzer alle zwei Jahre statt. Auch für Nutzer des MZ und der I<sub>A</sub>BS haben entsprechende Veranstaltungen bereits stattgefunden (siehe Lüttinger 1999 für Mikrozensus und den Sonderband der Mitteilungen aus der Arbeitsmarkt- und Berufsforschung in 1997). Über die Nutzergruppe des IAB-Betriebspanels liegen keine Angaben vor. Hauptnutzer scheinen die Mitarbeiter des IAB zu sein. HIS Daten werden ebenfalls hauptsächlich von Mitarbeitern des HIS und des IAB ge-

nutzt. Der Nutzerkreis der HIS-Daten ist ebenfalls weitgehend auf HIS Mitarbeiter beschränkt.

Das SOEP hat mittlerweile auch einen relativ großen Kreis von Nutzern, deren Arbeitsort im Ausland liegt. Dies hat auch erheblich zur intensiven Nutzung dieser Datenbasis beigetragen. Aufgrund der erwähnten rechtlichen Regelungen beschränkt sich die Nutzung der Mikrodaten der Amtlichen Statistik auf Inländer.

## Literaturverzeichnis

- Abowd, J. M., und F. Kramarz (1999), The analysis of labor markets using matched employer-employee data: *Handbook of Labor Economics*, 3, S. 2629-2710.
- Alex, L. (2000), 20 Jahre BIBB/IAB-Erhebungen, *BWP* 2/2000.
- Alex, L. und H. Bau (Hrsg.) (1999), *Wandel beruflicher Anforderungen*. Bielefeld.
- Baltagi, B. H. und U. Blien (1998), The German wage curve: Evidence from the IAB employment sample, *Economics letters*, Bd. 61, 2, 135-142.
- Baumann, C. (1999), Self-employment as an indicator of assimilation?: Evidence from foreign entrepreneurs in Germany, in: P. Lüttinger (Hrsg.), *Sozialstrukturanalysen mit dem Mikrozensus*, S. 235-270.
- Bechtold, S. und J. Chlumsky (2000), Amtliche Statistik und sozio-oekonomische Fragestellungen, *Wirtschaft und Statistik*, 2, S. 81-86.
- Becker, I. (2000), Einkommensverteilung in Deutschland: Strukturanalyse der Ungleichheit nach Einkommenskomponenten, *Arbeitspapiere des EVS-Projekts „Personelle Einkommensverteilung in der Bundesrepublik Deutschland“*, 25, Frankfurt a. M.
- Bellmann, L. (1997), Das Betriebspanel des IAB, in: R. Hujer, U. Rendtel, und G. G. Wagner (Hrsg.), *Wirtschafts- und sozialwissenschaftliche Panel-Studien*, Sonderhefte zum Allgemeinen Statistischen Archiv, Heft 30, Göttingen, S. 169 - 182.
- Bellmann, B., S. Bender und T. Schank (1999), Flexibilität der Qualifikationsstruktur aus betrieblicher Sicht: Substitutionalität oder Komplementarität, *Jahrbücher für Nationalökonomie und Statistik*, 1/2, S. 109-126.
- Bellmann, L., A. Reinberg und M. Tessaring (1994), „Bildungsexpansion, Qualifikationsstruktur und Einkommensverteilung: eine Analyse mit Daten des Mikrozensus und der Beschäftigtenstatistik“, in: L. R. Lüdeke (Hg.), *Bildung, Bildungsfinanzierung und Einkommensverteilung*, 13-70, Duncker & Humblot, Berlin.
- Bellmann, L., und S. Kohaut (1999), Betriebliche Beschäftigungsentwicklung und Innovationsaktivitäten: Ergebnisse aus dem IAB-Betriebspanel 1998, *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 32, 4, S. 416-422.
- Bellmann, L., und M. Lahner (1998), Betriebliche Erwartungen zur Ausbildungs- und Beschäftigungsentwicklung in West- und Ostdeutschland: Ergebnisse des IAB-Betriebspanels 1997, *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 31, 1, S. 79-92.
- Bellmann, L. und T. Schank (2000), Innovations, wages and demand for heterogeneous labour: new evidence from a matched employer-employee data-set, *IZA Discussion paper* 112.
- Bellmann, L. und U. Blien (1997), Lohnhöhe, Arbeitslosigkeit und Betriebsgröße: Eine Längsschnittanalyse mit dem IAB-Betriebspanel 1993 – 1995, in: J. Kühl, M. Lahner und J. Wagner (Hrsg.). *Die Nachfrageseite des Arbeitsmarktes: Ergebnisse aus Analysen mit deutschen Firmenpaneldaten*, Beiträge zur Arbeitsmarkt- und Berufsforschung , 204, Nürnberg, S. 255-273.
- Bender, S., C. Dustmann, D. Margolis and C. Meghir (1999), Worker displacement in France and Germany. *Institute for Fiscal Studies, Working paper* 99-14.
- Bender, S., J. Hilzendegen, G. Rohwer und H. Rudolph (1996), *Die IAB-Beschäftigtenstichprobe 1975 - 1990: Eine praktische Einführung*. Nürnberg, IAB. Beiträge zur Arbeitsmarkt- und Berufsforschung.

- Bender, S., A. Haas und C. Klose (2000), IAB employment subsample 1975 - 1995: opportunities for analysis provided by the anonymised subsample, *IZA Discussion paper* 117.
- Blanke, K., M. Ehling und N. Schwarz (1996), Zeit im Blickfeld : Ergebnisse einer repräsentativen Zeitbudgeterhebung, Schriftenreihe des Bundesministeriums für Familie, Senioren, Frauen und Jugend 121, Stuttgart [u.a.].
- Brand, R., S. Bender und S. Kohaut (1999), Möglichkeiten der Erstellung eines Scientific-Use Files aus dem IAB-Betriebspanel, in: Statistisches Bundesamt (Hrsg.), *Möglichkeiten einer wissenschaftlichen Nutzung von Unternehmensdaten aus der amtlichen Statistik*, S. 148-167.
- Brückner, H. und K. U. Mayer (1995), Lebensverläufe und gesellschaftlicher Wandel. Konzeption, Design und Methodik der Erhebung von Lebensverläufen der Geburtsjahrgänge 1954-1956 und 1959-1961. Teile I-III, Materialien aus der Bildungsforschung Nr. 48, Berlin: Max-Planck-Institut für Bildungsforschung.
- Bulmahn, G. und M. Kräkel (1999), Das NRW-Firmenpanel, in: Statistisches Bundesamt (Hrsg.), *Moeglichkeiten einer wissenschaftlichen Nutzung von Unternehmensdaten aus der amtlichen Statistik*, S. 97-103.
- Bundesinstitut für Berufsbildung (1995), „Erwerb und Verwertung beruflicher Qualifikationen.“, *BIBB/IAB Erhebung 1991/92*. Erhebungs- und Datendokumentation für das Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln, BIBB, Berlin.
- Bundesanstalt für Arbeit, verschiedene Jahrgänge, *Strukturanalyse: Bestände sowie Zu- und Abgänge an Arbeitslosen und offenen Stellen*.
- Burmester, M. (2000), Sozialhilfebezug in Ost und West – Eine vergleichende Analyse auf Basis der 25% Stichprobe, *Wirtschaft und Statistik*, 5, S. 384-395.
- Cornelsen, C. (1995), „Erwerbstätige mit Hochschulabschluß nach Hauptfachrichtungen“, *Wirtschaft und Statistik*, 11, 810 - 818.
- Cramer, U. (1986), Zur Stabilität von Beschäftigung: Erste Ergebnisse d. IAB-Stichprobe aus der Beschäftigtenstatistik, *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 19, 2, S. 243-256.
- Deutscher Industrie- und Handelstag (DIHT), 1995, Politik für mehr Beschäftigung braucht bessere Arbeitsmarktstatistik, Positionspapier, Bonn.
- Dietrich, H. (1999), Empirische Befunde zur selbständigen Erwerbstätigkeit unter besonderer Beruecksichtigung scheinselfständiger Erwerbsverhältnisse, *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 32, 1, S. 85-101.
- Domsch, M., T. J. Gerpott, S. Haugrund und M. Merfort (1990), *Personalentwicklung in der Industrieforschung: Eine empirische Untersuchung in forschungsintensiven Großunternehmen zum Problem der Überalterung von Fachwissen*, Poeschel, Stuttgart.
- Düll, H. und P. Ellguth (1999), Atypische Beschäftigung : Arbeit ohne betriebliche Interessenvertretung? ; Empirische Analysen mit dem IAB-Betriebspanel zum Einfluss von Betriebsräten auf befristete und geringfügige Beschäftigung, *WSI-Mitteilungen*, 52, 3, S. 165-176.
- Entorf, H., M. Gollac und F. Kramarz (1999), New technologies, wages, and worker selection, *Journal of Labor Economics*, 17, 3, S. 464-491.
- Essig, H. (2000), Darstellung der Einkommen nach dem Europaeischen System Volkswirtschaftlicher Gesamtrechnungen, *Wirtschaft und Statistik*, 2, S. 87-97.
- Fleischer, M. (1998), Patenting and Industrial Performance: The Case of the Machine Tool Industry, *WZB discussion paper* FS IV 98-9.

- Franz, W. (1996), *Arbeitsmarktökonomik*, 3. Aufl., Springer, Berlin.
- Fritsch, N. (1997), Die Betriebsdatei der Beschäftigtenstatistik: Ansatz und Analysepotenziale, in: R. Hujer, R., U. Rendtel und G. G. Wagner (Hrsg), 1997, *Wirtschafts- und sozialwissenschaftliche Panel-Studien: Datenstrukturen und Analyseverfahren*, *Sonderhefte zum Allgemeinen Statistischen Archiv*, 30, S. 127-147.
- Gerlach, K., O. Hübler und W. Meyer (Hg.) (1998), *Ökonomische Analysen betrieblicher Strukturen und Entwicklungen. Das Hannoveraner Firmenpanel*, Campus, Frankfurt/New York.
- Gerlach, K. und J. Wagner (1992), Betriebsdaten der amtlichen Statistik - eine (fast) ungenutzte Informationsquelle, *Staatswissenschaften und Staatspraxis*, 3.Jg., Heft 4, 578 - 596.
- Gerlach, K. und J. Wagner (1993), Zur Arbeitsplatzdynamik in der hannoverschen Industrie von 1978 bis 1991 - Eine empirische Untersuchung mit Betriebsdaten, *Statistische Monatshefte Niedersachsen* 2/93, 53 - 62.
- Gilberg, R., D. Hess und H. Schroeder (1999), Wiedereingliederung von Langzeitarbeitslosen: Chancen und Risiken im Erwerbsverlauf, *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 32, 3, S. 281-299.
- Hahlen, J. (1999), Begrüßung, in: Statistisches Bundesamt (Hrsg.), *Kooperation zwischen Wissenschaft und amtlicher Statistik: Praxis und Perspektiven* ; Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden, S. 7-11.
- Hart, P. E. und A. Shipman (1991), Skill Shortages in Britain and Germany, *International Journal of Manpower*, 12, 4, S. 18-30.
- Hamermesh, D. S (1999), LEEping into the future of labor economics: the research potential of linking employer and employee data, *Labour Economics*, 6, 1, S. 25-41.
- Harabi, N. (1998), Innovations through vertical relations between firms, suppliers and customers: a study of German firms, *Industry and innovation*, 5, 2, S. 157-179.
- Hauser, R. (1999), Personelle Primär- und Sekundärverteilung der Einkommen unter dem Einfluss sich ändernder wirtschaftlicher und sozialpolitischer Rahmenbedingungen: eine empirische Analyse auf der Basis der Einkommens- und Verbrauchsstichproben 1973-1993, *Allgemeines Statistisches Archiv*, 83, 1, S. 88-110.
- Hauser, R. und E. Becker (1999), Changes in the distribution of pre-government and post-government income in Germany 1973 – 1993, *Arbeitspapiere des EVS-Projekts "Personelle Einkommensverteilung in der Bundesrepublik Deutschland*, Nr. 20. Frankfurt a. M.
- Hauser, R. und E. Becker (1999), Wird unsere Einkommensverteilung immer ungleicher? Einige Forschungsergebnisse, In: D. Döring (Hrsg.) *Sozialstaat in der Globalisierung* S. 89-116.
- Hauser, R., G. G. Wagner und K. F. Zimmermann (1998), Erfolgsbedingungen empirischer Wirtschaftsforschung und empirisch gestützter wirtschafts- und sozialpolitischer Beratung, in: *Allgemeines Statistisches Archiv*, 82, 3, S. 369-379.
- Hujer, R., U. Rendtel und G. G. Wagner (1997), *Wirtschafts- und sozialwissenschaftliche Panel-Studien: Datenstrukturen und Analyseverfahren*, Sonderhefte zum Allgemeinen Statistischen Archiv, 30.
- Hübler, O. (1997), Der Arbeitsmarkt - Monitor des IAB, in: R. Hujer et al., *Wirtschafts- und sozialwissenschaftliche Panel-Studien: Datenstrukturen und Analyseverfahren*, S. 149-167.

- Hübler, O. (1995), Arbeitsplatzsuche Ostdeutscher in Westdeutschland zur Bedeutung myopischer und permanenter Individual- und Globaleffekte bei Paneldaten, *Ifo-Studien*, 41, 3, S. 381-408.
- Inkmann, J., W. Pohlmeier und L. A. Ricci (1998), Where to patent? : Theory and evidence on international patenting, *ZEW Discussion paper* 98-35.
- Jansen, R. (Hrsg.), (1993), *Qualifikation und Erwerbssituation im geeinten Deutschland*: Ein Überblick über die Ergebnisse der BIBB/IAB-Erhebung 1991/92.
- Jansen, R. (2000), *Auswirkungen des Strukturwandels auf die Arbeitsplätze*, Ergebnisse aus der BIBB/IAB-Erhebung 1998/99 zu Erwerb und Verwertung beruflicher Qualifikation, in BWP 2/2000.
- Kaiser, U. (1999), Die ZEW/Creditreform Konjunkturumfrage bei unternehmensnahen Dienstleistern, *Allgemeines Statistisches Archiv*, 83, 4, S. 447-451.
- Karr, W. (1999), Bereitstellung von Daten des IAB für die Wissenschaft, In: Statistisches Bundesamt (Hrsg.). *Kooperation zwischen Wissenschaft und amtlicher Statistik : Praxis und Perspektiven* ; Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden, S. 202-205
- Klös, H.-P. (1999), Die deutsche Arbeitsmarktstatistik, Aussagekraft und ihre Grenzen, in *IW-trends*, 1, S. 23-45. <http://www.iw-koeln.de/trends/tr1-99/tr1-5-09.htm>
- Knödler, H., (2000), Die Entwicklung der Kostenstruktur im Produzierenden Gewerbe seit 1950, *Wirtschaft und Statistik*, 3, S. 207-213.
- Kohaut, S. und C. Schnabel (1998), Flächentarifvertrag im Westen sehr viel weiter verbreitet als im Osten: Ergebnisse aus dem IAB-Betriebspanel, *IAB-Kurzbericht*, 1998-19.
- Köhler, S., B. Schimpl-Neimans und N. Schwarz (2000), Pilotprojekt zur Erleichterung der Nutzungsmöglichkeiten von faktisch anonymisierten Mikrodaten, *Wirtschaft und Statistik*, 1, S. 30-37.
- Kraft, K. (1991), The incentive effects of dismissals, efficiency wages, piece-rates and profit-sharing, *Review of Economics and Statistics*, 73, 3, S. 451-459.
- Krais, B. und L. Trommer (1995), „Akademiker-Beschäftigung: Sonderauswertung aus der Volkszählung 1987,“ *Materialien aus der Bildungsforschung* 54, Max-Planck-Institut für Bildungsforschung, Berlin.
- Lehner, F., M. Baethge, J. Kühl und F. Stille (Hg.) (1998), *Beschäftigung durch Innovation. Eine Literaturstudie*, Rainer Hamp Verlag, München.
- Lewin, K., K.-H. Minks und S. Uhde (1996), Abitur - Berufsausbildung - Studium : zur Strategie der Doppelqualifizierung von Abiturienten, *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 29, 3, S. 431-454.
- Lüttinger, P. (Hrsg.), (1999), *Sozialstrukturanalysen mit dem Mikrozensus, Konferenz "Forschung mit dem Mikrozensus: Analysen zur Sozialstruktur und zum Arbeitsmarkt"* , Oktober 1998 in Mannheim, ZUMA.
- Manser, M. E. (1998), Existing labor market data: current and potential research uses, in: J. C. Haltiwanger (Hrsg), *Labor statistics measurement issues*, S. 9-46
- Mavromaras, K. und H. Rudolph (1998), Temporary separations and firm size in the German labour market, *Oxford Bulletin of Economics and Statistics*, 60, 2, S. 215-225.
- Minks, K.-H. (1995), „Berufs- und Einsatzprofile von Fachhochschul- und Universitätsingenieuren“, *HIS-Kurzinformation A 13/94*, Hochschul-Informationssystem-GmbH (Hg.) Hannover.

- Minks, K.-H. (1998), Das Potenzial für Selbständigkeit unter Hochschulabsolventen. Sonderauswertung aus den Absolventenbefragungen der Prüfungsjahrgänge 1989 und 1993, *HIS-Kurzinformation A 1/98*, Hochschul-Informationssystem-GmbH (Hg.) Hannover.
- Möller, J. (1999), Die Entwicklung der qualifikatorischen Lohn- und Beschäftigungsstruktur in Deutschland: eine empirische Bestandsaufnahme, *Jahrbücher für Nationalökonomie und Statistik*, 1/2, S. 8-31.
- Müller, W. (1999), Amtliche Statistik und empirische Forschung, in: Statistisches Bundesamt (Hrsg.), *Kooperation zwischen Wissenschaft und amtlicher Statistik: Praxis und Perspektiven*; Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden, S. 17 - 35.
- Neubauer, W. (1999), Überlegungen zu einem innovativen Datenmanagement der amtlichen Statistik, in: Statistisches Bundesamt (Hrsg.), *Kooperation zwischen Wissenschaft und amtlicher Statistik: Praxis und Perspektiven*; Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden, S. 98-106.
- Nowack, M., (1997), Das Internet als Herausforderung für die amtliche Statistik, *Allgemeines Statistisches Archiv*, 81, S. 417-428.
- Pfeiffer, F. (1999), Existenzgründerpotenziale unter Arbeitsuchenden: empirische Evidenz auf der Basis des Mikrozensus, *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 32, 3, S. 300-314.
- Pfeiffer, F. und M. Falk (1999), *Der Faktor Humankapital in der Volkswirtschaft*, ZEW Wirtschaftsanalysen, Band 35, Nomos, Baden-Baden.
- Pfeiffer, F. und F. Reize (2000), Business start-ups by the unemployed: an econometric analysis based on firm data, erscheint in *Labour Economics*.
- Schares, C. (1998), Anforderungen der Wirtschaftspolitik an die Arbeitsmarktstatistik - Erfahrungen aus der wirtschaftspolitischen Beratung, in: J. Schupp et al. (Hrsg.), *Arbeitsmarktstatistik zwischen Realität und Fiktion*, S. 17-29.
- Schimpl-Neimanns, B. (2000), Weitere Mikrozensen fuer die Forschung leichter nutzbar, *Allgemeines statistisches Archiv*, 84, 1, S. 95.
- Schüler, K. (1987), Demographischer Bezugsrahmen zur Einkommensverteilung nach Haushaltsgruppen in den Volkswirtschaftlichen Gesamtrechnungen, *Wirtschaft und Statistik*, S. 361-370.
- Schüler, K. (1990), Verfügbares Einkommen nach Haushaltsgruppen in erweiterter Haushaltsgliederung 1972 bis 1988, *Wirtschaft und Statistik*, S. 182-194.
- Schnabel, C. und J. Wagner (1992), Unions and innovation: Evidence from German micro data, *Economics letters*, 39, 3, S. 369-373.
- Schoer, K., (1990), Erwerbstätige 1970 bis 1989: Quellen für die Fortschreibung und Rückrechnung, *Wirtschaft und Statistik*, 4, S. 247 - 251.
- Schupp, J., F. Büchel, M. Diewald und R. Habich (Hrsg.) (1998): *Arbeitsmarktstatistik zwischen Realität und Fiktion*.
- Schupp, J. (1999), Zur Erhebungsproblematik geringfügiger Beschäftigung: Ein Strukturvergleich des Mikrozensus mit dem sozio-oekonomischen Panel und dem Europäischen Haushaltspanel, in: P. Lüttinger (Hrsg.), *Sozialstrukturanalysen mit dem Mikrozensus*, S. 93-118. ZUMA Mannheim.
- Schwarz, N. (1996), Ehrenamtliches Engagement in Deutschland: Ergebnisse der Zeitbudgeterhebung 1991/92, *Wirtschaft und Statistik*, S. 259-266.

- Smolny, W. und T. Schneeweis (1999), Innovation, Wachstum und Beschaeftigung : eine empirische Untersuchung auf der Basis des ifo Unternehmenspanels, *Jahrbuecher fuer Nationaloekonomie und Statistik*, 3/4, S. 453-472.
- Statistisches Bundesamt (1999a), *Kooperation zwischen Wissenschaft und amtlicher Statistik: Praxis und Perspektiven*, Beiträge zum Symposium am 31. Mai/1. Juni 1999, in Wiesbaden.
- Statistisches Bundesamt (1999b), *Möglichkeiten einer wissenschaftlichen Nutzung von Unternehmensdaten aus der amtlichen Statistik*, Schriftenreihe Spektrum Bundesstatistik 14.
- Statistisches Bundesamt (versch. Jgg.), Bevölkerung und Erwerbstätigkeit, Fachserie 1, Reihe 4. 1. 2. *Beruf, Ausbildung, und Arbeitsbedingungen der Erwerbstätigten*, Stuttgart.
- Statistisches Bundesamt (versch. Jgg.), Bevölkerung und Erwerbstätigkeit, Fachserie 1, Reihe 4. 2. 1, *Struktur der sozialversicherungspflichtig Beschäftigten*, Stuttgart.
- Statistisches Bundesamt (1997), *Fünf Jahre Deutscher Mikrozensus*, Internet: [www.statistikbund.de/presse/deutsch/pm/p6229032.htm](http://www.statistikbund.de/presse/deutsch/pm/p6229032.htm).
- Statistisches Bundesamt, Sozialeleistungen, Fachserie 13, Leistungen an Asylbewerber, Reihe 7, Stuttgart.
- Statistisches Bundesamt, Sozialeleistungen, Fachserie 13, Sozialhilfe, Reihe 7, Stuttgart.
- Statistisches Bundesamt, Fachserie 15, *Wirtschaftsrechnungen*. Reihe 1, Einnahmen und Ausgaben ausgewählter privater Haushalte, Stuttgart.
- Statistisches Bundesamt, Fachserie 4, *Produzierendes Gewerbe*, Reihe 4.1.1, Beschäftigung, Umsatz und Energieversorgung der Unternehmen und Betriebe im Bergbau und im Verarbeitenden Gewerbe.
- Statistisches Bundesamt, Fachserie 2, *Unternehmen und Arbeitsstätten*, Reihe 1.5.1.
- Statistisches Bundesamt, Fachserie 2, *Unternehmen und Arbeitsstätten*, Reihe 1.5.2,.
- Statistisches Bundesamt, Fachserie 2, *Unternehmen und Arbeitsstätten*, Reihe 1.6.1.
- Statistisches Bundesamt, Fachserie 2, *Unternehmen und Arbeitsstätten*, Reihe 1.6.2.
- Statistisches Bundesamt, Fachserie 4, *Produzierendes Gewerbe*, Reihe 4.3.1, Kostenstruktur der Unternehmen im Bergbau, Grundstoff- und Produktionsgütergewerbe.
- Statistisches Bundesamt, Fachserie 4, *Produzierendes Gewerbe*, Reihe 4.3.2, Kostenstruktur der Unternehmen im Investitionsgüterproduzierenden Gewerbe.
- Statistisches Bundesamt, Fachserie 4, *Produzierendes Gewerbe*, Reihe 4.3.3, Kostenstruktur der Unternehmen im Verbrauchsgüterproduzierenden Gewerbe und im Nahrungs- und Genußmittelgewerbe.
- Statistisches Bundesamt, Fachserie 16, *Löhne und Gehälter - Arbeitskostenerhebung*, Heft 2, Arbeitskosten im Groß- und Einzelhandel sowie im Bank- und Versicherungsgewerbe.
- Statistisches Bundesamt, Fachserie 16, *Löhne und Gehälter - Arbeitskostenerhebung*, Heft 1, Arbeitskosten im Produzierenden Gewerbe, 1988.
- Statistisches Landesamt Baden-Württemberg, LIS, Benutzerhandbuch, Stand Januar 1993.
- Stephan, G. (1999), Die niedersächsische Gehalts- und Lohnstrukturerhebung als Basis arbeitsökonomischer Auswertungen, in: Statistisches Bundesamt (Hrsg.), *Möglichkeiten einer wissenschaftlichen Nutzung von Unternehmensdaten aus der amtlichen Statistik*, S. 104-117.
- Stephan, G. (2000), *Firm wage Differentials in Germany*, Institut für Quantitative Wirtschaftsforschung, Hannover, mimeo.

- Steiner, V. und K. Wagner (1998), Has earnings inequality in Germany changed in the 1980's?, *Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 118, 1, S. 29-59.
- Steiner, V. und F. Kraus (1995), Haben Teilnehmer an Arbeitsbeschaffungsmassnahmen in Ostdeutschland bessere Wiederbeschäftigungschancen als Arbeitslose?, in: V. Steiner (Hrsg.), *Mikroökonomik des Arbeitsmarktes*, S. 387-423
- Stifterverband Statistik (1996), *Forschung und Entwicklung in der Wirtschaft*, Stifterverband für die Deutsche Wissenschaft, Essen.
- Statistischen Beirat (1999), *Empfehlungen zur Weiterentwicklung der amtlichen Statistik, Bericht des Statistischen Beirats an die Bundesregierung*, Wiesbaden, Juni 1999, [http://www.statistik-bund.de/presse/deutsch/pm/beirat\\_1.htm](http://www.statistik-bund.de/presse/deutsch/pm/beirat_1.htm)
- Strotmann, H. (1999), Zur wissenschaftlichen Nutzung von Betriebsdaten aus der amtlichen Statistik: erste Erfahrungen und Ergebnisse eines Kooperationsprojektes zwischen dem Statistischen Landesamt Baden-Württemberg und der Universität Hohenheim, in: *Möglichkeiten einer wissenschaftlichen Nutzung von Unternehmensdaten aus der amtlichen Statistik*, S. 29-44.
- Stubig, H.-J. (1999), Das Niedrigeinkommenpanel (NIEP) als Beispiel für Ressortforschung, in: Statistisches Bundesamt (Hrsg.), *Kooperation zwischen Wissenschaft und amtlicher Statistik: Praxis und Perspektiven* ; Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden, 52 – 58.
- Wagner, G. G, R. V. Burkhauser und F. Behringer (1993), The English language public use file of the German Socio-Economic Panel, *Journal of Human Resources*, 28, 2, 429-440.
- Wagner, J. (1999), Nutzung von betrieblichen Einzeldaten aus der amtlichen Statistik durch externe Wissenschaftlicher: Modelle, Erfahrungen, Perspektiven, in: *Möglichkeiten einer wissenschaftlichen Nutzung von Unternehmensdaten aus der amtlichen Statistik*, S. 9-17.
- Wiegert, R. (1998), Zur Gegenwart und Zukunft der amtlichen Statistik in Deutschland, *Jahrbucher für Nationalökonomie und Statistik*, 217, 1, S. 138-150.
- Wiener, B. (1995), *Arbeitsmarktdaten Sachsen-Anhalt: Wegweiser zu den Arbeitsmarktdaten Sachsen-Anhalt*, Ministerium für Arbeit, Soziales und Gesundheit des Landes Sachsen-Anhalt (Hrsg.), Forschungsbeiträge zum Arbeitsmarkt in Sachsen-Anhalt, 8.
- Weisshuhn, G. (1998), Kritik am bestehenden Instrumentarium der Arbeitsmarktstatistik, in: Schupp, J. (Hrsg), *Arbeitsmarktstatistik zwischen Realität und Fiktion*, S. 31-39.
- Widmaier, U. und P. Hauptmanns (1997), Das NIFA-Panel im deutschen Maschinenbau - ein Überblick, in: R. Hujer et al., *Wirtschafts- und sozialwissenschaftliche Panelstudien. Datenstrukturen und Analyseverfahren*, S. 229-241.
- Wolf, E. (1999), Dynamik der Arbeitszeitstruktur: Welche Rolle spielt der Strukturwandel?, in: P. Lüttinger (Hrsg.), *Sozialstrukturanalysen mit dem Mikrozensus*, S. 119-148.
- Zwick, M. (1999), Steuerstatistische Einzeldaten und ihre Auswertungsmöglichkeiten fuer die Wissenschaft, *Allgemeines Statistisches Archiv*, 83, 2, S. 248-253.
- Ziegler, H. (1999), Bessere Datengrundlagen für die Wissenschaft ergeben bessere Handlungsgrundlagen für die Politik; in: Statistisches Bundesamt (Hrsg.), *Kooperation zwischen Wissenschaft und amtlicher Statistik: Praxis und Perspektiven* ; Beiträge zum Symposium am 31. Mai/1. Juni 1999 in Wiesbaden, S. 12 – 16.

## Anhang

Nur ein Teil aller verfügbaren Mikrodatsätze im Bereich Erwerbstätigkeit und Einkommen fällt unter die hier verwendete Abgrenzungskriterien (Wiederholungsbefragung, keine regionale Begrenzung, Repräsentativität). Tabelle 5 enthält eine Übersicht über Datensätze, die nicht unter diese Kriterien fallen, für Forschungszwecke aber bereits Verwendung gefunden haben. Allerdings musste auch hier eine Auswahl getroffen werden. Für Datensätze, die die Grundgesamtheit der Bevölkerung oder Betriebe bzw. Unternehmen abbilden, ist eine Stichprobe von mindestens 3000 Personen oder Haushalten bzw. 1.500 Betrieben oder Unternehmen erforderlich. Für einzelne Subpopulationen (einzelne Branchen oder bestimmte Personengruppen) sind auch weniger Beobachtungen zulässig. Die Übersicht erhebt keinen Anspruch auf Vollständigkeit.<sup>35</sup> Die Übersicht enthält auch Datensätze, die sehr unregelmäßig wiederholt werden, wie z. B. die Volkszählung oder Arbeitsstättenzählung, sowie Datensätze deren Erhebungen mittlerweile eingestellt sind (z.B. Arbeitsmarktmonitor für Ostdeutschland, IFO Unternehmenspanel).

### Personen- und Haushaltsdaten

Der Arbeitsmarktmonitor des IAB ist eine achtwellige Panelbefragung der ostdeutschen Bevölkerung im erwerbsfähigen Alter im Zeitraum 1990 bis 1994 (siehe Hübler 1994) und wird nicht weiter fortgeführt. Frageschwerpunkt ist die Erwerbstätigkeit, Suchverhalten und Wirkung arbeitsmarktpolitischer Maßnahmen. Auf Basis des Arbeitsmarktmonitors sind eine Vielzahl wissenschaftlicher Arbeiten entstanden (siehe die Zusammenfassung in Hübler 1997). Unter anderem liegen Untersuchungen zur Entlohnung (Steiner und Bellmann 1995), Wiederbeschäftigungschancen von Teilnehmern an ABM (Steiner und Kraus 1995) und Arbeitsplatzsuche (Hübler 1995) vor.

Die Luxembourg Income Study (LIS) und Luxembourg Employment Study (LES) gehen auf eine Initiative der Regierung Luxemburgs und des Center for Population, Poverty, and Policy Studies (CEPS) Anfang der achtziger Jahre zurück und werden mittlerweile von 24 fördernden Institutionen aus 18 Ländern finanziell unterstützt. Auf deutscher Seite hat diese Aufgabe ZUMA Mannheim übernommen (vgl. Smeeding 1999). Neben den meisten Ländern der EU beteiligen sich auch Institutionen aus Osteuropa, Amerika, Australien und Asien. Das Hauptziel von LIS besteht in der Sammlung, Aufbereitung („Harmonisierung“) und Weitergabe von für Einkommensanalysen relevanter Mikrodaten für wissenschaftliche Zwecke. Durch „Harmonisierung“ der von den nationalen Stellen an LIS weitergegebene Daten sollen vergleichende Analysen ermöglicht werden. Für Deutschland sind zur Zeit die Mikrodaten des SOEP für die Jahre 1984, 1989 und 1994, der Erhebungen 1973, 1978 und 1983 der unten beschriebenen Einkommens- und Verbrauchsstichprobe des Statistischen Bundesamts und einer einmaligen Befragung (die „Transferumfrage 1981“) des ehemaligen von den Universitäten Frankfurt/Main und Mannheim durchgeführten SFB3 der DFG verfügbar. Gegenüber dem ECHP besteht ein wesentlicher Nachteil in den völlig unterschiedlichen Erhebungskon-

---

<sup>35</sup> Beispielsweise wurde auf die Auflistung von Absolventenbefragungen (siehe Bargel), Befragung Privater Arbeitsvermittlung (Dincher und Gaugler 1999), Ad-hoc-Umfrage zur Arbeitsmarktlage im Dienstleistungssektor (EU 1999) verzichtet.

zepten und Erhebungsjahren der einzelnen nationalen Mikrodatensätze. Hingegen besteht ein wesentlicher Vorteil von LIS in der Möglichkeit des Fernrechnens. Seit 1998 können in SAS, SPSS und STATA geschriebene Auswertungsprogramme ohne zusätzliche Kosten über email bearbeitet werden.

Die IAB-Selbstständigenstudie ist eine Querschnittsbefragung zu Selbstständigkeit für ca. 1.000 Selbstständige (siehe Dietrich 1999). Angaben über Nutzerkreis und Zugangsmöglichkeiten stehen nicht zur Verfügung.

## **Firmendaten**

Das IFO Unternehmenspanel, Mannheimer Innovationspanel und NIFA Panel wurden nicht in die Analyse einbezogen, da bei allen Befragungen das Innovationsverhalten im Vordergrund steht. Das Hannoveraner Firmenpanel bezieht sich nur auf Betriebe im niedersächsischen Verarbeitenden Gewerbe.

Schwerpunkt des IFO Unternehmenspanels ist das Innovationsverhalten von Unternehmen. Für Beschäftigung liegen jedoch nur eingeschränkte Informationen vor. Ursprünglicher Auftraggeber war das BMBF. Im Rahmen eines Pilotprojekts Ende der 90er Jahre (siehe Schneeweiss und Smolny 1997) wurde das IFO Unternehmenspanels auch externen Nutzern zur Verfügung gestellt. Mit den IFO Unternehmensdaten gibt es ebenfalls eine Reihe von Veröffentlichungen (siehe Rottman und Ruchinsky 1998 und Smolny 1999).

Das NIFA-Panel ist eine regelmäßige repräsentative Befragung von Betrieben in der Maschinenbaubranche (siehe ausführlich Widmaier und Hauptmanns 1997). Auftraggeber der Befragung ist die Universität Bochum. Das NIFA wurde bis 1996 im Rahmen des DFG-Sonderforschungsbereichs 187 "Neue Informationstechnologien und flexible Arbeitssysteme" an der Ruhr-Universität Bochum und ab 1996 als DFG-Einzelprojekt finanziert. Schwerpunkte der Befragung sind die Diffusion von CIM-Technologien, Konzepte und Formen der Arbeitsorganisation, sowie Personalstruktur und Qualifikation in der Maschinenbaufertigung. Es gibt eine Reihe von detaillierten Fragen zur Erwerbstätigkeit (Mitarbeiterzahl in einzelnen Produktionsbereichen und deren Qualifikationsgrad, Arbeitszeitregelung, Personalprobleme in verschiedenen Mitarbeitergruppen, Maßnahmen gegen Personalengpässe). Der Datensatz wird nach Prüf- und Korrekturarbeiten sehr zügig an das Zentralarchiv weitergegeben. Zur Erschwerung der Anonymisierung sind alle Informationen zum Betriebsort entfernt worden. Zudem sind die Angaben zu Umsatz und Beschäftigung zu Klassen zusammengefaßt worden. Zur Zeit können beim Zentralarchiv die ersten acht Wellen für den Zeitraum 1991-1998 in leicht anonymisierter Form bezogen werden. Trotz der zügigen Weitergabe der Daten ist das NIFA-Panel von externen Nutzern kaum verwendet worden (Fleischer 1998). Eine Erklärung hierfür ist, dass Umsatz und Beschäftigte nicht mehr als metrische Größe vorhanden sind und dass die Betrachtung einer einzelnen Branche für wissenschaftliche Untersuchungen oft wenig attraktiv ist.

Das Hannoveraner Firmenpanel ist eine repräsentative Befragung von Betrieben des Verarbeitenden Gewerbes in Niedersachsen. Schwerpunkte der Befragung sind Entlohnung, Arbeitsorganisation, Beschäftigung, Arbeitszeit und Innovationsverhalten. Insgesamt sind vier Befragungen für den Zeitraum 1994-1997 verfügbar. Zugangsmöglichkeiten bestehen im Rechnen „vor Ort“. Es liegen bereits eine Reihe von wissenschaftlichen Untersuchungen vor

(für einen Überblick vgl. Gerlach, Hübler und Meyer 1998). Es sind auch bereits erste Untersuchungen von externen Wissenschaftlern, die das Rechnen „vor Ort“ genutzt haben, verfügbar (siehe z. B. Peters 2000). Die Erstellung eines anonymisierten Datensatzes ist geplant.

Seit 1993 führt das Zentrum für Europäische Wirtschaftsforschung (ZEW) jährlich systematische Erhebungen zum Innovationsverhalten der deutschen Wirtschaft durch. In Zusammenarbeit mit infas - Institut für angewandte Sozialwissenschaft (infas) wird eine für das Produzierende Gewerbe repräsentative Datenbasis aufgebaut und gepflegt (Mannheimer Innovationspanel, MIP). Parallel wird seit 1995 eine entsprechende Erhebung auch im Dienstleistungssektor (MIP-DL) durchgeführt. Allerdings sind zu Beschäftigung und Qualifikation nur wenige Fragen verfügbar. An der Erhebung im Produzierenden Gewerbe beteiligen sich jährlich etwa 2500 Unternehmen. Dabei wird versucht, die gleichen Unternehmen in die jährlichen Befragungen einzubeziehen (Panelbefragung), um die Entwicklung auf der Unternehmensebene verfolgen zu können. Um das Panel aktuell zu halten, werden alle zwei Jahre neugegründete Unternehmen in den Stichprobenrahmen aufgenommen. Auf Basis der MIP-Daten liegen eine Reihe von Untersuchungen interner und externer Wissenschaftler vor (zu den Arbeiten externer Nutzer siehe beispielsweise Harabi 1998, Inkmann et al. 1998).

Seit 1998 erhebt das BIBB Angaben über aktuelle Themen der betrieblichen Berufsausbildung für ca. 1.700 Ausbildungsbetriebe und Institutionen.<sup>36</sup> Erste Ergebnisse der Referenz-Betriebs-System (RBS) Umfrage sind in Alex und Höcke (2000) publiziert. Über Verfügbarkeit, Bereitstellung und Weitergabe liegen noch keine Angaben vor. Die BIBB Befragung von Betrieben und Auszubildenden zu den neuen IT-Berufen wird erst in diesem Jahr abgeschlossen werden. Nähere Angaben liegen ebenfalls nicht vor. Über die BIBB-Erhebung Ausbildungsverhalten neu gegründeter Betriebe liegen ebenfalls keine Angaben über Verfügbarkeit vor.

---

<sup>36</sup> Siehe hierzu [http://www.bibb.de/forum/projekte/rbs/was\\_ist.htm](http://www.bibb.de/forum/projekte/rbs/was_ist.htm).



Tabelle 5: Weitere Datensätze zu Erwerbstätigkeit und Einkommen

Datensatz	Auftraggeber	Erhebungszeitraum	Fallzahl pro Welle	Gebiet	Stichprobe	Fragenschwerpunkte	Weitergabe
Alterssicherung in Deutschland	BMAS	1986	16000	D	Zufallsauswahl der Bevölkerung ab 55 J.; Ziehung: Melderegister der Einwohnermeldeämter.	Erwerbs- und Lebensverlauf, Einkommenssituation und Alterssicherung älterer Personen.	Ja, auf Antrag beim ZA
Arbeitsmarktmonitor	IAB	1990-1994	7500-10500	NBL	0,1% der ostdeutschen Bevölkerung im erwerbsfähigen Alter	Erwerbstätigkeit, Arbeitsmarktpolitische Maßnahmen	Ja, auf Antrag beim ZA
Arbeitsmarktmonitor Sachsen-Anhalt	Land Sachsen-Anhalt	1990-1997	6775	Sachsen-Anhalt	Repräsentativstichprobe der Bevölkerung	Erwerbstätigkeit, Arbeitsmarktpolitische Maßnahmen	Ja, auf Antrag beim ZA
Arbeitsstättenzählung	Bund	1987	>200 Tsd.	D	Vollerhebung	Anzahl der Beschäftigten und die Betriebliche Stellung	Keine Angaben
Ausbildungsverhalten neu gegründeter Betriebe <sup>e</sup>	BIBB	1998	2100	D	Repräsentative Befragung von Betrieben in der privaten Wirtschaft	Ausbildungsbereitschaft neuer Betriebe	Keine Angaben
BIBB IT Studie	BIBB	1999	5000	D	Befragung von Betrieben und Auszubildenden	Neue IT-Berufe	Keine Angaben
Caritas-Armutsuntersuchung	Prof. Hauser (Universität Frankfurt)	1990	3200	D	Repräsentativ-Befragung der Caritas Klienten	Lebenslagen und Armutsbetroffenheit der Klienten der Deutschen Caritas.	Ja, auf Antrag beim ZA
Erwerbswünsche und Erwerbsverhalten von Frauen	IAB	1991, 1995	ca. 7000	D	geschichtete Zufallsauswahl von Frauen zwischen 18 und 65	berufliche Situation bzw. Erwerbswünsche	Ja, auf Antrag beim ZA
Hannoveraner Firmenpanel	Universität Hannover	1994-1997	1500	Niedersachsen	Unternehmensregister Niedersachsen, nur Verarbeit. Gewerbe	Entlohnung, Organisation, Beschäftigung, Innovation	Vor Ort Rechnen
IAB-Selbständigenumfrage	IAB	1997	1000	D	Keine Angaben	Selbstständigkeit	Keine Angaben

Tabelle 5: Fortsetzung

Datensatz	Auftraggeber	Erhebungs- zeitraum	Fallzahl pro Welle	Gebiet	Stichprobe	Fragenschwerpunkte	Weitergabe
IAB-Selbständigen- umfrage	IAB	1997	1000	D	Keine Angaben	Selbstständigkeit	Keine Angaben
Luxembourg Income Study (LES) <sup>b</sup>	Regierung Luxem- burg, CEPS	SOEP 84,89,94; EVS: 83	10000	Industrie- länder	Wie SOEP und EVS	Wie SOEP und EVS	Fernrechnen SPSS, STATA
Luxembourg Employ- ment Study (LES) <sup>b</sup>	Regierung Luxem- burg, CEPS			s. o.	Wie SOEP und EVS	Wie SOEP und EVS	S . o
Mannheimer Innovati- onspanel (MIP) <sup>d</sup>	BMBF, ZEW	1993-an- dauernd, ab 95 Dienstleistun- gen	3000 VG, 2500 DL	D	Geschichtete Stichprobe aus den ZEW-Creditreform Daten	Schwerpunkt Innovationsverhal- ten	Ja,. faktisch anony. Daten, Rechnen „vor Ort“
NIEP (Niedrigein- kommenspanel)	BMAS	1998 – 2000 6 Wellen	ca. 2000 Haushalte	D	repräsent. Stichprobe von Haus- halte mit deutschsprachigen Haushaltsvorstand oder Partner	Einkommen im Niedriglohnbe- reich	Keine Angaben
NIFA (Neue Informati- onstechnologien und flexible Arbeitssyste- me) <sup>a</sup>	Universität Bo- chum	1991-1998	ca. 1600	D	Betriebsdatei der Beschäftigten- statistik; Betriebe in Maschinen- bau mit mehr als 20 Beschäftigte,	Diffusion von CIM- Technologien, Konzepte und Formen der Arbeitsorganisation, Qualifikation, Organisation	Ja, auf Antrag beim ZA
Referenz-Betriebs- System (RBS) <sup>c</sup>	BIBB	1998	ca. 1700	D	Keine Angaben	betriebliche Berufsausbildung	Keine Angaben
Soll und Haben	Spiegel, Manager Magazin	1980, 1984, 1989,1994	8700	D	Wohnbevölkerung ab 14 Jahren	Einstellung zum Geld und Ver- halten bei der Anlage von Geld.	Ja, auf Antrag beim ZA
Volkszählung	Bund	1987	>20 Mio	D	Totalerhebung	Wie Mikrozensus	Keine Angaben

<sup>a</sup> Weitere Informationen siehe <http://www.pw2.ruhr-uni-bochum.de/projekte/nifa/>; <sup>b</sup> Weitere Informationen siehe <http://www.lis.ceps.lu/index.htm>; <sup>c</sup> siehe [http://www.bibb.de/forum/projekte/rbs/was\\_ist.htm](http://www.bibb.de/forum/projekte/rbs/was_ist.htm); <sup>d</sup> siehe [http://www.zew.de/projekte/industrioeconomik/laufende\\_p/io02.html](http://www.zew.de/projekte/industrioeconomik/laufende_p/io02.html); <sup>e</sup> [http://www.bibb.de/publikat/bwp/5\\_98/aktuell.htm](http://www.bibb.de/publikat/bwp/5_98/aktuell.htm); <sup>f</sup> siehe <http://www.biat.uni-flensburg.de/bibb-it/Auftrag/Auftrag-Inhalte.htm>.

Quelle: Eigene Zusammenstellung.