

Mehr und bessere Daten für die Arbeitsmarktforschung: Vorschläge zur effizienten Erschließung von bislang ungenutztem Potenzial der amtlichen Statistik

Schneider, Hilmar

Veröffentlichungsversion / Published Version

Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

SSG Sozialwissenschaften, USB Köln

Empfohlene Zitierung / Suggested Citation:

Schneider, H. (2009). *Mehr und bessere Daten für die Arbeitsmarktforschung: Vorschläge zur effizienten Erschließung von bislang ungenutztem Potenzial der amtlichen Statistik*. (RatSWD Working Paper Series, 106). Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-418053>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.



Rat für Sozial- und
Wirtschaftsdaten (RatSWD)

www.ratswd.de

RatSWD

Working Paper Series

Working Paper

No. 106

Mehr und bessere Daten für die Arbeitsmarktforschung

Vorschläge zur effizienten Erschließung von bislang
ungenutztem Potenzial der amtlichen Statistik

Hilmar Schneider

Juli 2009

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienen empirischen Forschungsarbeiten, sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/ 2008 Heike Solga; 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

Mehr und bessere Daten für die Arbeitsmarktforschung
Vorschläge zur effizienten Erschließung von bislang ungenutztem Potenzial der
amtlichen Statistik

Hilmar Schneider

(IZA)

1. Einleitung

Die ökonomische Beurteilung von Vorgängen auf dem Arbeitsmarkt erfordert eine solide und aussagekräftige Datengrundlage. Eine Kerngröße zu deren Verständnis bilden Stundenlöhne. Stundenlöhne erklären beispielsweise, ob und in welchem Umfang Menschen arbeiten und konsumieren, warum und wie lange sie arbeitslos bleiben und wie viel sie in Bildung investieren. Stundenlöhne erklären darüber hinaus, in welchem Umfang Firmen Menschen entlassen oder einstellen, für welchen Standort sie sich entscheiden und welche Form der Spezialisierung für sie optimal ist. Nicht zuletzt sind Stundenlöhne von erheblicher Bedeutung für die Beantwortung von Verteilungsfragen. Umso erstaunlicher mutet es daher an, dass seitens der amtlichen Statistik für diesen zentralen Bereich der Wirtschaft bislang keine hinreichende Datengrundlage bereitgestellt wird. Das ließe sich mit relativ geringem Aufwand ändern. Dazu bedarf es jedoch letztlich der Unterstützung durch den Gesetzgeber. Dies betrifft zum einen die Konzeption amtlicher Erhebungen und zum anderen die Schaffung einer gesetzlichen Grundlage für Verknüpfung von amtlichen Mikrodaten zu wissenschaftlichen Zwecken.

Nach wie vor entspringt die Logik der durchgeführten Erhebungen einem antiquierten Verständnis von Statistik, das seine Bestimmung darin sieht, Aggregatzahlen mit Hilfe eng zweckgebundener und unabhängiger Stichproben zu gewinnen. So werden inhaltlich zusammengehörige Merkmale wie Arbeitszeit und Entlohnung unabhängig voneinander in getrennten Erhebungen erfasst. Stundenlöhne lassen sich dadurch nur im Aggregat berechnen. Das analytische Potenzial solcher Datensätze bleibt solange von begrenztem Wert, wie keine breite Verknüpfung von strukturellen Merkmalen auf der Ebene von Beobachtungseinheiten möglich ist.

Wären die Stichproben sogar so konzipiert, dass eine Verknüpfung von Informationen auf Betriebsebene mit Informationen auf der Personen- und Haushaltsebene möglich wäre, könnte das Analysepotenzial der Daten erheblich ausgeweitet werden bei gleichzeitiger Reduktion des Erhebungsaufwands. Nicht zuletzt deshalb hat die Kommission zur Verbesserung der informationellen Infrastruktur den Gesetzgeber dazu aufgefordert, die Möglichkeit exakter Datenverknüpfung – ohne ausdrückliche Zustimmung aller einzelnen Befragten – für rein statistische Zwecke gesetzlich zu regeln (vgl. KVI 2001, Empfehlung 34).

Aggregierte Statistiken reichen für eine solide Analyse nicht aus, weil sie keine eindeutigen Rückschlüsse auf die strukturellen Ursachen von Unterschieden oder Veränderungen im Aggregat zulassen. In der wissenschaftlichen Literatur ist dieses

Phänomen als ökologischer Fehlschluss bekannt (Robinson 1950). Typische Fragen der Arbeitsmarktforschung lassen sich nur auf der Basis von Mikrodaten beantworten. Die Heterogenität von Beobachtungseinheiten auf der Mikroebene – für die Arbeitsmarktforschung kommen hier vor allem Individuen, Haushalte und Firmen in Betracht – dient in diesem Zusammenhang als identifizierendes Moment. Dabei sind zwei Arten der Heterogenität relevant: Heterogenität zwischen Beobachtungseinheiten auf der Mikroebene (Between Variance) und Heterogenität über die Zeit (Within Variance). Handelt es sich bei dem zu untersuchenden Zusammenhang um einen stationären Prozess und sind die erklärungsrelevanten Merkmale beobachtbar, genügt die Erhebung von Querschnittsdaten, d.h. die Erhebung von Heterogenität zwischen Beobachtungseinheiten. Die darüber hinausgehende Erhebung in Form von Paneldaten ist zum einen dann relevant, wenn Kohorteneffekte auftreten, zum anderen dann, wenn die Beobachtbarkeit der erklärungsrelevanten Merkmale eingeschränkt ist, was grundsätzlich nie ausgeschlossen werden kann. Mit Hilfe von Paneldaten lässt sich der möglicherweise verzerrende Einfluss unbeobachteter Heterogenität neutralisieren (vgl. Baltagi 1995). Daraus folgt ein grundsätzliches Erfordernis zur Erhebung von Paneldaten auf Mikroebene. Eine Vielzahl entsprechender Daten wird bereits heute im Rahmen der amtlichen Statistik erhoben oder fällt im Rahmen von Verwaltungsprozessen an.

Das damit verbundene Anliegen ist im Übrigen nicht völlig neu. In einer Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik im Bereich Erwerbstätigkeit und Einkommen haben sich Falk/Steiner (2000) mit den angesprochenen Problemen auseinandergesetzt und eine Reihe von Empfehlungen formuliert. Der vorliegende Beitrag greift diese Empfehlungen auf und bilanziert die seither erzielten Fortschritte. Darauf aufbauend wird der verbleibende Handlungsbedarf herausgearbeitet. Im Vordergrund steht hier der Vorschlag einer koordinierten Stichprobenziehung, die eine Verknüpfung von personenbezogenen und betriebsbezogenen Informationen zuließe und als Weiterentwicklung der Empfehlungen von Falk/Steiner zu betrachten ist.

2. Paradebeispiel Stundenlöhne

Für ökonomische Analysen des Arbeitsmarktgeschehens sind in erster Linie die faktischen Stundenlöhne relevant. Dazu bedarf es sowohl einer Information über die gezahlten Löhne als auch einer Information über die dafür tatsächlich geleistete Arbeitszeit. In der amtlichen Statistik werden beide Informationen jedoch nur in voneinander getrennten Erhebungen

erfasst und das zudem in unzureichender Form. So wird beispielsweise im Mikrozensus zwar die geleistete Arbeitszeit erhoben, nicht aber das dabei erzielte Erwerbseinkommen. Stattdessen wird lediglich das persönliche Nettoeinkommen bzw. das Haushaltsnettoeinkommen erfragt. Da in diesen Größen auch Transfers und andere Einkommen in unbekannter Höhe einfließen können, ist eine Berechnung von Stundenlöhnen nicht möglich. Etwa ein Sechstel der deutschen Haushalte bezieht Transfers wie Wohngeld und Arbeitslosenunterstützung (vgl. Rudolph 2008). Knapp ein Viertel aller Haushalte bezieht nach den Angaben des Statistischen Bundesamtes Kindergeld oder Erziehungsgeld. Insbesondere für Bezieher von Niedriglöhnen sorgt das Steuer- und Transfersystem für eine starke Einkommenskompensation (BMAS 2008), so dass eine approximative Verwendung der Nettoeinkommen anstelle von Erwerbseinkommen zu völlig irreführenden Schlussfolgerungen führen dürfte.

In der Einkommens- und Verbrauchsstichprobe (EVS) wird seit 2003 neben detaillierten Einkommenskategorien immerhin die wöchentliche Arbeitszeit erfasst, wenn auch nur die vertraglich vereinbarte. Da in das erzielte Einkommen aber auch Überstunden einfließen, werden die tatsächlichen Stundenlöhne bei einer Division von Erwerbseinkommen durch die vertraglich vereinbarte Arbeitszeit überschätzt. Hinzu kommt, dass die EVS nur alle fünf Jahre durchgeführt wird und auf bestimmte Haushaltstypen beschränkt ist. Außerdem handelt es sich bei der EVS um eine freiwillige Erhebung, was wiederum mit systematischen Selektionsverzerrungen einhergehen kann, die die Repräsentativität der Erhebung zusätzlich beeinträchtigen (vgl. Hauser 2009).

Auch die 1991/92 sowie 2001/02 durchgeführte Zeitbudgeterhebung überwindet die genannten Defizite nicht. Zwar enthält die Zeitbudgeterhebung eine detaillierte Aufzeichnung der Zeitverwendung im Tagesverlauf, doch lässt sich diese nicht mit dem erzielten Erwerbseinkommen in Verbindung bringen.

Eine direkte Erfassung von Stundenlöhnen durch die amtliche Statistik erfolgt lediglich auf der Betriebsebene. Dabei handelt es sich allerdings wiederum nur um vertraglich vereinbarte Stundenlöhne und erlaubt somit nur eingeschränkte Schlussfolgerungen im Hinblick auf die tatsächlich anfallenden Arbeitskosten für ein Unternehmen. Dies erlaubt keinerlei Verknüpfung mit Haushaltsmerkmalen und nur eine äußerst eingeschränkte Verknüpfung mit Personenmerkmalen. Selbst wenn eine Verknüpfung möglich wäre, ist zu berücksichtigen, dass auf der Firmenseite nur Bruttostundenlöhne erhoben werden. Für Haushalte ist dagegen der Nettostundenlohn die entscheidende Größe.

3. Die Empfehlungen von Falk/Steiner

Die amtliche Statistik verfügt über eine Reihe von Mikrodatsätzen, die für die Arbeitsmarktforschung von Relevanz sind. Falk/Steiner (2000) betrachten in diesem Zusammenhang folgende Datensätze:

1. Europäisches Haushaltspanel
2. Mikrozensus
3. Einkommens- und Verbrauchsstichprobe
4. Zeitbudgeterhebung
5. Gehalts- und Lohnstrukturerhebung
6. Amtliche Industriestatistik für Betriebe und Unternehmen
7. Sozialhilfestatistik
8. Kostenstrukturerhebung

Ihre Expertise kommt zu insgesamt 15 Empfehlungen, von denen 8 im Hinblick auf diese Datensätze relevant sind. Diese sind im Folgenden stichwortartig zusammengefasst:

1. Weiterentwicklung des Mikrozensus zu einem arbeitsmarktrelevanten Datensatz durch Aufnahme von Einkommensmerkmalen
2. Weiterentwicklung der Einkommens- und Verbrauchsstichprobe als Instrument für Jahresrechnungen und Erweiterung um Arbeitszeitmerkmale
3. Unterjährige Erhebung des Mikrozensus
4. Weiterentwicklung des Mikrozensus als Access Panel
5. Öffnung des Datenzugangs für interessierte Wissenschaftler
6. Bereitstellung personenbezogener Mikrodaten der amtlichen Statistik in faktisch anonymisierter Form
7. Bereitstellung von Betriebs- und Unternehmensdaten der amtlichen Statistik in einfach anonymisierter Form
8. Vergleichende Eignungsprüfung des Datenzugangs über datenschutzrechtlich geschützte Rechenzentren und Datenfernrechnen im Rahmen eines Pilotprojekts

Die beiden erstgenannten Empfehlungen sind als essentiell zu betrachten und beziehen sich auf das eingangs geschilderte Problem, dass die vorhandenen Mikrodaten von der Arbeitsmarktforschung bislang wenig genutzt werden, weil zentrale Variablen nicht vorhanden sind. Im Bereich personenbezogener Daten gilt dies vor allem für den Stundenlohn, im Bereich firmenbezogener Daten gilt dies für die Zerlegung der Wertschöpfung in Arbeit und Kapital.

4. Fortschrittsbilanz seit 2000

Zunächst einmal ist positiv festzustellen, dass die Zahl der verfügbaren Mikrodatensätze, die für die Arbeitsmarktforschung von Interesse sein können, mit der Einrichtung der Forschungsdatenzentren beim Statistischen Bundesamt und den Statistischen Landesämtern erheblich zugenommen hat. Nahezu alle Datenerhebungen, die amtlicherseits durchgeführt werden, sind inzwischen entweder als Scientific-Use-Files oder als Onsite-Files für interessierte Wissenschaftler zugänglich. Den Forderungen nach einer Öffnung des Datenzugangs wurde damit in starkem Maße Rechnung getragen. Bislang unzureichend – zumindest was den Bereich der amtlichen Statistik anbelangt – ist lediglich die Umsetzung des kontrollierten Fernrechnens, obwohl es bereits im KVI-Gutachten eine wichtige Rolle spielte und aus Sicht der Kommission einen integralen Bestandteil der Arbeit der neu gegründeten Forschungsdatenzentren darstellt (vgl. KVI 2001, Empfehlung 29).

Neben den Forschungsdatenzentren haben sich ebenfalls auf Empfehlung der KVI zwei Datenservicezentren etabliert, deren Aufgabe in der Bereitstellung von speziellen Dienstleistungen für Datennutzer und Datenproduzenten besteht (vgl. Schneider/Wolf 2008). Das bei der GESIS angesiedelte German Microdata Lab hat es sich zur Aufgabe gemacht, besondere Dokumentations- und Nutzungshilfen für amtlichen Mikrodaten zu erarbeiten. Das am Forschungsinstitut zur Zukunft (IZA) beheimatete Internationale Datenservicezentrum (IDSC) richtet sich an eine internationale Forschergemeinschaft und stellt einerseits Metadaten über arbeitsmarktrelevante Mikrodatensätze zur Verfügung und unterstützt andererseits Forscher aus dem Ausland beim Zugang zu deutschen Mikrodaten. Hier wurde unter anderem ein Tool für kontrolliertes Fernrechnen entwickelt, das unter der Bezeichnung JoSuA (Job Submission Application) firmiert und sich als Prototyp für den Einsatz in den Forschungsdatenzentren eignet. Bislang kommt es jedoch lediglich für ausländische Wissenschaftler zum Einsatz, die im Rahmen eines Kooperationsprojekts mit dem IZA Scientific-Use-Files der amtlichen Statistik nutzen möchten. Die Forscher senden ihre Auswertungsanfragen via JoSuA an das IDSC, von wo aus der eigentliche Datenzugriff nach einer datenschutzrechtlichen Prüfung des Inputs durch das IDSC erfolgt. Nach einer datenschutzrechtlichen Prüfung des Outputs durch das IDSC wird dieser – soweit zulässig – an den anfragenden Forscher zurückgeschickt.

Auch wenn die Zahl der amtlicherseits zur Verfügung gestellten Mikrodatensätze erheblich zugenommen hat, ist Quantität hier nicht gleichbedeutend mit Qualität. Der Mehrwert etwa von betrieblichen Daten zur Weiterbildung im Rahmen der Europäischen Erhebung zur beruflichen Weiterbildung (CVTS) hält sich in Grenzen, da im Rahmen dieser

Erhebung weder individualspezifische Merkmale erfasst werden, noch eine Verknüpfung mit vorhandenen Individualdatensätzen möglich ist. Ähnliches gilt für die Statistik der Studenten oder die Statistik der Eheschließungen, um nur zwei weitere Beispiele zu nennen.

Im Hinblick auf die beiden erstgenannten Forderungen von Falk/Steiner ist der erzielte Fortschritt als bescheiden einzustufen. Mit der Neufassung des Mikrozensusgesetzes 2005 wurde die Chance zur Erweiterung um das Merkmal Erwerbseinkommen auf absehbare Zeit vertan. Immerhin aber wird in der EVS seit 2003 die vertraglich vereinbarte wöchentliche Arbeitszeit erfragt. Bis dahin beschränkte sich die Erhebung auf die Kategorien Teilzeit- und Vollzeitbeschäftigung sowie geringfügige Beschäftigung. Wie bereits angesprochen, ist die Eignung der EVS zur Berechnung von Stundenlöhnen damit allerdings dennoch keineswegs zufriedenstellend gelöst.

Der Forderung nach einer unterjährigen Erhebung des Mikrozensus wurde mit der Neufassung des Mikrozensusgesetzes Rechnung getragen. Bis dahin bezogen sich die Angaben des Mikrozensus stets auf einen Stichtag im Mai des jeweiligen Erhebungsjahres. Der Mikrozensus war daher in der Vergangenheit nicht dazu geeignet, die erheblichen saisonalen Schwankungen des Erwerbsverhaltens abzubilden. Nach umfangreichen Testerhebungen in den Jahren 2003 und 2004 wird seit Januar 2005 eine unterjährliche Erfassung in monatlichen Intervallen durchgeführt. Die Testerhebungen haben der Arbeitsmarktforschung in Gestalt der Piloterhebung zum ILO-Erwerbsstatus immerhin eine interessante neue Mikrodatenbasis beschert.

Der Vorschlag, den Mikrozensus als so genanntes Access Panel weiterzuentwickeln wurde seitens der amtlichen Statistik ebenfalls aufgegriffen. Seit dem Jahr 2004 werden die Teilnehmer des Mikrozensus um ihr Einverständnis gebeten, sich auf freiwilliger Basis an weiteren Stichproben der amtlichen Statistik zu beteiligen (vgl. Körner et al. 2006). Die Vorteile des Verfahrens bestehen darin, sehr flexibel Zufallsstichproben ziehen zu können, höhere Ausschöpfungsquoten realisieren zu können als dies mit herkömmlichen Verfahren der Stichprobenziehung möglich ist, vor allem aber darin, vorhandene Informationen nicht mehrfach erheben zu müssen. Potenziell könnte sich somit das konstatierte Defizit der getrennten Erhebung sachlich zusammengehörender Merkmale überwinden lassen. De facto wird dieses Potenzial bislang jedoch nicht annähernd ausgeschöpft, da nur ein relativ geringer Teil der Merkmale, die im Mikrozensus erhoben werden, in die entsprechende Masterdatei übernommen wird. So findet sich von den Merkmalen zur Arbeitszeit lediglich die normalerweise geleistete Arbeitszeit je Woche in der Masterdatei. Die Verknüpfung etwa mit

der EVS bietet somit keinen Informationsgewinn im Hinblick auf die Generierung von Stundenlöhnen.

Die auf dem Mikrozensus basierende Auswahl von potenziellen Befragungsteilnehmern dient vor allem zur Stichprobenauswahl für die Gemeinschaftsstatistik über Einkommen und Lebensbedingungen (EU-SILC). Diese 2005 zum erstenmal durchgeführte Panel-Erhebung hat die Nachfolge des bis 2001 erhobenen europäischen Haushaltspanels angetreten. Leider wurde auch bei dieser Erhebung die Chance versäumt, eine solide Datenbasis zur Erfassung individueller Stundenlöhne zu schaffen. Zwar werden detaillierte Informationen zur aktuellen Arbeitszeit erhoben, doch bezüglich des Einkommens beschränkt sich die Erhebung auf das monatliche Nettoeinkommen, ohne nach Einkommenskomponenten zu differenzieren. Eine differenzierte Einkommenserhebung erfolgt lediglich retrospektiv für das jeweils zurückliegende Jahr. Dem steht jedoch keine entsprechende Arbeitszeiterfassung gegenüber. Denkbar wäre zwar ein zeitversetzter Abgleich der aktuellen Arbeitszeitinformationen mit den retrospektiv erhobenen Erwerbseinkommenseingaben des jeweiligen Folgejahres. Dies lässt sich allerdings nur für Personen mit einer durchgehenden Beschäftigung durchführen. Bei Personen mit Erwerbsunterbrechungen ist hingegen keine eindeutige Zuordnung zwischen den retrospektiven Jahreseinkommensangaben und den aktuellen Arbeitszeitangaben möglich.

Auch wenn im Hinblick auf die Generierung von individuellen Stundenlöhnen auf der Basis amtlicher Mikrodaten wenig Fortschritt erzielt worden ist, ist aktuell eine Entwicklung zu beobachten, die Anlass zu Hoffnung gibt. Hier handelt es sich um Ansätze, die die Idee zur Verknüpfung von Mikrodaten aus verschiedenen Quellen aufgreifen und im Rahmen der bestehenden gesetzlichen Möglichkeiten so weit wie möglich umsetzen. Dass dabei mitunter sogar Institutionengrenzen überwunden werden, ist zweifellos ein Verdienst der geschaffenen Dateninfrastruktur. An erster Stelle ist in diesem Zusammenhang das Projekt KombiFiD zu nennen (vgl. Bender/Möller 2009, J. Wagner 2009).¹ Hier geht es um die Zusammenführung von Firmendaten von Statistischem Bundesamt, Bundesbank und Bundesagentur für Arbeit (BA) mit dem Ziel der Erstellung eines kombinierten Firmendatensatzes.

In der koordinierten Verknüpfung bislang unverbundener Erhebungen steckt zweifellos ein gewaltiges Potenzial für die Weiterentwicklung der amtlichen Statistik. Ein wichtiger Schritt in diese Richtung ist auch in der Konzeption des geplanten Zensus 2011 zu sehen, bei dem die verfügbaren Informationen aus diversen Registern auf der Ebene von Individuen bzw. Haushalten zusammengeführt werden sollen (vgl. Statistische Ämter des Bundes und der

¹ siehe auch: <http://www.kombifid.de>

Länder 2004; Heinzel 2006). Das dazu entwickelte Verfahren dürfte wichtige Impulse zu einer Neukonzeption von koordinierten Stichprobenziehungen in der amtlichen Statistik liefern

5. Resultierender bzw. verbleibender Handlungsbedarf

Aus dem Vergleich zwischen den dargestellten Datenanforderungen und dem inzwischen Erreichten ergeben sich im Wesentlichen vier Handlungsempfehlungen, die sich stichpunktartig wie folgt zusammenfassen lassen:

- Umsetzung des kontrollierten Fernrechnens in den Forschungsdatenzentren der amtlichen Statistik
- Erweiterung des Befragungsprogramms von Haushaltsstatistiken im Hinblick auf Stundenlöhne
- Gesetzliche Regelung exakter Datenverknüpfung – ohne ausdrückliche Zustimmung aller einzelnen Befragten – für rein statistische Zwecke
- Koordinierte Durchführung statistischer Erhebungen zur Schaffung von Datenverknüpfungsmöglichkeiten

Die Forderung nach einer Umsetzung von kontrolliertem Fernrechnen bedarf keiner weiteren Erläuterung. Die Forderung nach einer Erweiterung des Befragungsprogramms von Haushaltsstatistiken im Hinblick auf Stundenlöhne steht in einem Spannungsverhältnis zu den nachfolgenden Empfehlungen. Wäre es möglich, verschiedene Datenquellen miteinander zu verknüpfen, würde sich die Forderung nach einer Erweiterung des Befragungsprogramms erübrigen. Insofern ist diese Forderung als Second-Best-Lösung zu verstehen, denn sie löst nur ein spezifisches, wenn auch elementares, Problem, während die koordinierte Durchführung statistischer Erhebungen mit dem Ziel der Schaffung von Datenverknüpfungsmöglichkeiten nicht nur das Problem fehlender Stundenlöhne lösen würde, sondern zugleich viele weitere. Ohne eine vernünftige gesetzliche Grundlage ist dieses allerdings nicht zu erwarten.

Solange eine solche grundlegende gesetzliche Lösung nicht in Sicht ist, bleibt keine Alternative zu der Forderung nach einer Ausweitung des Erhebungsprogramms in Mikrozensus und EVS. Die nächste Chance dafür bietet sich im Zuge der auf europäischer Ebene angestrebten Harmonisierung der nationalen Haushaltsstatistiken. Die Geltungsdauer des Mikrozensusgesetzes endet mit dem Jahr 2012. Ein Jahr später läuft die gesetzliche Grundlage für die EVS aus. Es ist zu erwarten, dass der Gesetzgeber eine umfassende

Neukonzeption beider Statistiken beschließen wird, die sich an europäischen Vorgaben orientiert. Bereits im November 2007 fand dazu in Mannheim ein gemeinsam von GESIS dem Rat für Sozial- und Wirtschaftsdaten organisierter Workshop statt, dessen Ziel darin bestand, den Dialog zwischen Wissenschaft und amtlicher Statistik im Hinblick auf Erhebungsprogramme der amtlichen Haushaltsstatistik zu fördern. Das Statistische Bundesamt befasst sich mit der Frage im Rahmen des Projekts "Reform der Haushaltsstatistiken". Darüber hinaus wurde von Bund und Ländern inzwischen ein Lenkungsgremium berufen, um das weitere Vorgehen zu beraten. Ein abgestimmtes Konzept müsste spätestens Ende 2010 vorliegen, um zeitgerecht in das Gesetzgebungsverfahren eingetaktet werden zu können. Auf europäischer Ebene finden die entsprechenden Überlegungen in der Planung eines neuen European Household Survey (EHS) ihren Ausdruck. 2008 wurde bereits eine Piloterhebung durchgeführt.

Die Alternative der koordinierten Verknüpfung von Datenquellen erlaubt hingegen eine weitaus effektivere Verbreiterung der Informationsbasis, da sie einerseits die Qualität der erhobenen Daten verbessert und andererseits eine Reduzierung des Erhebungsaufwands zulässt. Dabei muss die Verknüpfungsmöglichkeit nicht auf Erhebungen der Statistischen Ämter beschränkt bleiben. Denkbar wäre beispielsweise auch eine Verknüpfung der Mikrozensusstichprobe mit Verwaltungsdaten der BA. Während die BA-Daten keine Informationen zur Arbeitszeit enthalten, fehlen im Mikrozensus die Informationen zum Erwerbseinkommen. Jeweils für sich genommen ist die Aussagefähigkeit beider Datensätze somit relativ eingeschränkt. Durch die Kombination von beiden würde dagegen eine mächtige Analysebasis entstehen, indem beispielsweise die in den BA-Daten fehlenden Informationen zur Arbeitszeit mit den im Mikrozensus erhobenen Informationen ergänzt würden.

Eine solche Strategie weist im Übrigen weit über die Empfehlungen von Falk/Steiner (2000) hinaus. Während sich diese auf die komplementäre Erhebung von personenbezogenen bzw. haushaltsbezogenen Daten innerhalb der amtlichen Statistik beschränkten, zeigt das Projekt KombiFiD die Möglichkeiten auf, die sich aus einer koordinierten Verknüpfung von Firmen- und Individual- bzw. Haushaltsdaten über institutionelle Grenzen hinweg ergeben. Bei KombiFiD ist seitens der BA zwar nur vorgesehen, aggregierte Individualdaten wie Anzahl der Mitarbeiter, Altersstruktur der Mitarbeiter etc. beizusteuern, doch vom Grundsatz her eröffnet sich hier auch die Möglichkeit einer umgekehrten Perspektive in Form der Verknüpfung von Individualdaten mit Firmendaten und damit die Erweiterung in Richtung eines Linked-Employer-Employee-Datensatzes.

Koordinierte Verknüpfung von Datenquellen erfordert in diesem Fall eine abgestimmte Stichprobenziehung aus Haushalten und Firmen. So ist es beispielsweise denkbar, das Mikrozensus sample nicht mehr ausschließlich als Bevölkerungsstichprobe zu generieren, sondern als eine Stichprobe, in die alle Beschäftigten aus einem Firmensample eingehen. Allerdings würde sich der Mikrozensus nicht ausschließlich auf dieser Basis ziehen lassen, da ansonsten Arbeitslose und Nicht-Erwerbstätige von der Erhebung ausgeschlossen wären. Sinnvoller dürfte deshalb der umgekehrte Fall sein, der darin bestünde, die Arbeitgeber der Mikrozensus-Befragten als Basis für ein Unternehmenssample heranzuziehen.

Auch wenn die hier aufgezeigte Perspektive eher langfristiger Natur ist, gibt es keinen Grund, sie als unrealistisch abzutun. Mit der Neukonzeption des Zensus 2011 als registergestütztem Zensus ist ein Weg aufgezeigt, wie die grundsätzlichen Hürden überwunden werden können. Es sollte nun darum gehen, diesen Weg auch für die Forschung nutzbar zu machen.

Literatur:

- Baltagi, B.H. (1995): *Econometric Analysis of Panel Data*. (Wiley) Chichester, New York, Brisbane, Toronto, Singapore
- Bender, S.; Möller, J. (2009): *Developing the Research Data Infrastructure for the Social and Behavioral Sciences in Germany and Beyond*. Erscheint in: Solga, H.; Wagner, G.G. (Hrsg.): *KVI Reloaded*
- Bundesministerium für Arbeit und Soziales (2008): *Lebenslagen in Deutschland – Der 3. Armuts- und Reichtumsbericht der Bundesregierung*. Bonn
- Falk, M.; Steiner, V. (2000): *Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik im Bereich Erwerbstätigkeit und Einkommen*. Mannheim
- Hauser, R. (2009): *Household income, poverty, and wealth*. Erscheint in: Solga, H.; Wagner, G.G. (Hrsg.): *KVI Reloaded*
- Heinzel, A. (2006): *Volkszählung 2001 – Deutschland bereitet sich auf den registergestützten Zensus vor*. *Berliner Statistik*, 7/2006: 321-328
- Körner, T.; Nimmergut, A.; Nökel, J.; Rohloff, S. (2006): *Die Dauerstichprobe befragungsbereiter Haushalte*. *Wirtschaft und Statistik*, 5/2006: 451-467
- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (2001): *Wege zu einer besseren informationellen Infrastruktur*. (Nomos) Baden-Baden
- Robinson, W.S. (1950): *Ecological Correlations and the Behavior of Individuals*. *American Sociological Review* 15: 351–357
- Rudolph, H. (2008): *Erwerbstätigkeit in Transferhaushalten - Arm trotz Arbeit*. *IAB Forum*, 2/2008: 34-39
- Schneider, H.; Wolf C. (2008): *Die Datenservicezentren als Teil der informationellen Infrastruktur*. In: Rolf, G.; Zwick, M.; Wagner, G.G. (Hrsg.): *Fortschritte der informationellen Infrastruktur in Deutschland*. (Nomos) Baden-Baden: 236-249
- Statistische Ämter des Bundes und der Länder (2004): *Ergebnisse des Zensus 2001*. *Wirtschaft und Statistik*, 8/2004: 813-833
- Wagner, J. (2009): *Improvements and Future Challenges for the Research Infrastructure in the Field Firm Level Data*. Erscheint in: Solga, H.; Wagner, G.G. (Hrsg.): *KVI Reloaded*