Wirt Sozialstat Archiv (2008) 2: 189–191
DOI 10.1007/s11943-008-0043-x

VORWORT

# Betriebs- und Unternehmensdaten im Längsschnitt – Neue Datenangebote und ihre Forschungspotenziale

**Stefan Bender · Martin Rosemann · Sylvia Zühlke · Markus Zwick**

Online: 18. September 2008
© Springer 2008

Der Zugang der Wissenschaft zu Mikrodaten der amtlichen Statistik wurde in den letzten Jahren deutlich verbessert und mit der Einrichtung der Forschungsdatenzentren institutionalisiert. Zugleich entstanden im Bereich der Wirtschaftsstatistik neue Datenangebote. Mit dem Projekt „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ wurden wesentliche Grundlagen für die Bereitstellung von Scientific-Use-Files gelegt. Einzelne Wirtschaftsstatistiken können bereits als Scientific-Use-Files genutzt werden.

Um das Datenangebot im Bereich wirtschaftsstatistischer Paneldaten zu verbessern, wurde das Projekt „Wirtschaftsstatistische Paneldaten und faktische Anonymisierung“ ins Leben gerufen, das vom Bundesministerium für Bildung und Forschung (BMBF) auf Empfehlung des Rats für Sozial- und Wirtschaftsdaten gefördert wird. Gegenstand der Projektarbeiten sind die Erweiterung des Datenangebots im

S. Bender (✉)
Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB), Regensburger Str. 104, 90478 Nürnberg, Deutschland
e-mail: stefan.bender@iab.de

M. Rosemann
Institut für Angewandte Wirtschaftsforschung (IAW), Ob dem Himmelreich 1, 72074 Tübingen, Deutschland
e-mail: martin.rosemann@iaw.edu

S. Zühlke
Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen (LDS NRW), Roßstr. 76, 40476 Düsseldorf, Deutschland
e-mail: sylvia.zuehlke@lds.nrw.de

M. Zwick
Forschungsdatenzentrum des Statistischen Bundesamtes, Gustav-Stresemann-Ring 11, 65180 Wiesbaden, Deutschland
e-mail: markus.zwick@destatis.de

Bereich der wirtschaftsstatistischen Einzeldaten durch Längsschnittverknüpfungen von bisher vorrangig im Querschnitt ausgewerteten Statistiken und die Optimierung des Analysepotenzials wirtschaftsstatistischer Paneldaten. Darüber hinaus soll nach Möglichkeiten zur faktischen Anonymisierung von wirtschaftsstatistischen Paneldaten gesucht werden mit dem Ziel, diese auch als Scientific-Use-Files anbieten zu können.

Ziel des Projekts ist es, den Datenzugang zu Paneldaten der amtlichen Statistik und des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) bzw. der Bundesagentur für Arbeit (BA) weiter zu verbessern. Dabei werden verschiedene Zugangswege zu Einzeldatensätzen berücksichtigt: Das Fernrechnen, der direkte Zugang an einem Gastwissenschaftlerarbeitsplatz im Forschungsdatenzentrum und – der Schwerpunkt des Projekts – die Scientific-Use-Files zur Nutzung am eigenen Arbeitsplatz.

In einem ersten Schritt wurden bisher vorrangig im Querschnitt verwendete wirtschaftsstatistische Einzeldaten der statistischen Ämter zu Längsschnittdaten verknüpft. Damit wird das Datenangebot der Forschungsdatenzentren deutlich erweitert.

Auf der Konferenz „Betriebs- und Unternehmensdaten im Längsschnitt – Neue Datenangebote und ihre Forschungspotenziale“, die am 14. und 15. Juni 2007 in Tübingen stattfand, wurde über die laufenden Projektarbeiten berichtet. Insbesondere wurde über neue Datenangebote der amtlichen Statistik und ihre Forschungspotenziale informiert. Daneben wurde ein Ausblick gegeben auf die Problematik der Anonymisierung von wirtschaftsstatistischen Paneldaten und die geplante Verknüpfung von Statistiken im Rahmen der beiden Projekte „Amtliche Firmendaten in Deutschland (AFiD)“ und „Kombinierte Firmendaten in Deutschland (KombiFiD)“.

Der vorliegende Band des *AStA Wirtschafts- und Sozialstatistisches Archiv* dokumentiert ausgewählte Beiträge dieser Nutzerkonferenz.

Eingeleitet wird der Band durch einen Aufsatz von Maurice Brandt, Dirk Oberschachtsiek und Ramona Pohl mit dem Titel „Neue Datenangebote in den Forschungsdatenzentren – Betriebs- und Unternehmensdaten im Längsschnitt“. Die Autoren geben einen Überblick über das aktuelle wirtschaftsstatistische Datenangebot der Forschungsdatenzentren. Zudem liefern sie einen Ausblick auf die beiden Projekte AFiD und KombiFiD.

In den folgenden beiden Beiträgen stehen die Forschungspotenziale unterschiedlicher Datensätze im Vordergrund. Joachim Wagner stellt in seinem Beitrag „Die Forschungspotenziale der Betriebspaneldaten des Monatsberichts im Verarbeitenden Gewerbe“ anhand von drei inhaltlichen Fragestellungen dar: Zunächst geht er der weit verbreiteten These nach, dass der Mittelstand der entscheidende Jobmotor in Deutschland sei. Anschließend beschäftigt er sich mit der Frage, wie die Produktivität einerseits mit Marktein- und -austritten sowie andererseits mit der Exporttätigkeit zusammenhängt.

Nicole Guertzgen stellt „Das Forschungspotenzial von Linked Employer-Employee Daten am Beispiel von Lohneffekten der Tarifbindung“ dar. Im Gegensatz zu den meisten bekannten Studien basiert die Analyse auf einem Längsschnitt-Datensatz, dem Linked-Employer-Employee-Datensatz des IAB. Damit wird versucht, der Selektion von Individuen und Betrieben aufgrund unbeobachteter Merkmale in die Tarifbindung Rechnung zu tragen.

In den drei weiteren Beiträgen werden unterschiedliche Aspekten der Anonymisierung von wirtschaftsstatistischen Paneldaten thematisiert.

Zunächst beschäftigt sich der Beitrag „Risk assessment methodology for longitudinal business microdata“ von Rainer Lenz mit der Messung von Re-Identifikationsrisiken bei wirtschaftsstatistischen Paneldaten. Dabei wird berücksichtigt, dass einem potenziellen Datenangreifer im Fall von Paneldaten zusätzliche Informationen über die Entwicklung im Zeitablauf und über die Zusammenhänge zwischen unterschiedlichen Wellen zur Verfügung stehen.

Im Beitrag von Jörg Höhne mit dem Titel „Anonymisierungsverfahren für Paneldaten“ werden unterschiedliche Verfahren zur faktischen Anonymisierung wirtschaftsstatistischer Paneldaten dargestellt. Dabei liegt der Schwerpunkt auf datenverändernden Anonymisierungsverfahren, die sich bereits für die Anonymisierung wirtschaftsstatistischer Querschnittsdaten bewährt haben. In seinem Beitrag werden insbesondere Weiterentwicklungen der Verfahren vorgestellt.

Gegenstand des abschliessenden Beitrags „Within-Schätzung bei anonymisierten Paneldaten“ von Elena Biewen sind zwei spezielle datenverändernde Verfahren, das Individual Ranking und die multiplikative stochastische Überlagerung. Es wird untersucht, welche Auswirkungen diese Verfahren auf die Schätzung panelökonometrischer Modelle mit Hilfe des Within-Schätzers haben.

Die in diesem Band dokumentierten Konferenzbeiträge zeigen zum einen, welch bedeutende Forschungspotenziale in den Unternehmens- und Betriebsdaten der amtlichen Statistik liegen und wie wichtig gerade die Längsschnittverknüpfung dieser Daten für die Forschung sein kann. Zum anderen werden die methodischen Herausforderungen dargestellt, die mit der faktischen Anonymisierung von wirtschaftsstatistischen Paneldaten verbunden sind und welche Lösungsansätze es hierfür gibt. Die Beiträge verdeutlichen den Stellenwert des Projektes „Wirtschaftsstatistische Paneldaten und faktische Anonymisierung“ für die Weiterentwicklung des Datenangebots über Unternehmen und Betriebe in Deutschland. Zugleich wird bereits der Bogen zu neuen Projekten geschlagen, in deren Rahmen die Forschungsdatenzentren der statistischen Ämter des Bundes und der Länder sowie der BA ihre Dienstleistungsangebote weiterentwickeln werden. Das Analysepotenzial wirtschaftsstatistischer Einzeldaten wird hier noch einmal dadurch deutlich verbessert, dass unterschiedliche Wirtschaftsstatistiken – auch über die Grenzen der Datenanbieter hinweg – verknüpft werden.

Die in diesem Heft präsentierten Beiträge zeigen, wie eng und fruchtbar die Zusammenarbeit zwischen Datenanbietern und Datennutzern mittlerweile geworden ist.

**Danksagung** Die Herausgeber bedanken sich bei allen Gutachtern für ihre Beteiligung am Begutachtungsverfahren.

Düsseldorf/Nürnberg/Tübingen/Wiesbaden, Juli 2008

Stefan Bender,
Martin Rosemann,
Sylvia Zühlke,
Markus Zwick