

FDZ-Arbeitspapier  
Nr. 18

Hans-Peter Hafner  
Rainer Lenz



**STATISTISCHE ÄMTER**  
DES BUNDES UND DER LÄNDER  
FORSCHUNGSDATENZENTREN

Die Gehalts- und Lohn-  
strukturerhebung:  
Methodik, Datenzugang  
und Forschungspotential

2007

FDZ-Arbeitspapier  
Nr. 18

Hans-Peter Hafner  
Rainer Lenz



**STATISTISCHE ÄMTER**  
**DES BUNDES UND DER LÄNDER**  
**FORSCHUNGSDATENZENTREN**

Die Gehalts- und Lohn-  
strukturerhebung:  
Methodik, Datenzugang  
und Forschungspotential

2007

**Herausgeber:** Statistische Ämter des Bundes und der Länder  
**Herstellung:** Statistisches Bundesamt

**Fachliche Informationen**  
zu dieser Veröffentlichung:

Statistisches Bundesamt  
Forschungsdatenzentrum  
Tel.: 0611 75-4220  
Fax: 0611 72-4000  
forschungsdatenzentrum@destatis.de

Erscheinungsfolge: unregelmäßig  
Erschienen im Mai 2007

**Informationen zum Datenangebot:**

Statistisches Bundesamt  
Forschungsdatenzentrum  
Tel.: 0611 75-4220  
Fax: 0611 72-4000  
forschungsdatenzentrum@destatis.de

Forschungsdatenzentrum der  
Statistischen Landesämter  
– Geschäftsstelle –  
Tel.: 0211 9449-2876  
Fax: 0211 9449-2901  
forschungsdatenzentrum@lds.nrw.de

Diese Publikation wird kostenlos als **PDF-Datei** zum Download unter [www.forschungsdatenzentrum.de](http://www.forschungsdatenzentrum.de) angeboten.

© Statistisches Bundesamt, Wiesbaden 2007  
(im Auftrag der Herausgebergemeinschaft)

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

**Bei den enthaltenen statistischen Angaben handelt es sich um eigene Arbeitsergebnisse des genannten Autors im Zusammenhang mit der Nutzung der bereitgestellten Daten der Forschungsdatenzentren.**

# Die Gehalts- und Lohnstrukturerhebung: Methodik, Datenzugang und Forschungspotenzial

*Hans-Peter Hafner<sup>1</sup> und Rainer Lenz<sup>2</sup>*

## 1. Einleitung

Linked Employer Employee – Datensätze (LEED) ermöglichen es der Arbeitsmarktforschung, beobachtete Effekte in einen durch das Unternehmen verursachten und einen vom Beschäftigten abhängigen Anteil aufzuspalten. Seit Mitte der 1990er entwickelte sich die Anzahl der Analysen in diesem Bereich sprunghaft. Abowd und Kramarz 1999 geben einen Überblick über die in den 90ern bearbeiteten Fragestellungen und zu diesem Zeitpunkt verfügbaren Datensätze aus 17 Ländern.

Für Deutschland sind aktuell für die Wissenschaft verfügbar der LEED des IAB (LIAB) und die Gehalts- und Lohnstrukturerhebung der Statistischen Ämter des Bundes und der Länder. Eine Beschreibung des LIAB und ausgewählter mit ihm durchgeführter Studien findet sich in Alda, Bender und Gartner 2005.

In diesem Aufsatz stellen wir die Gehalts- und Lohnstrukturerhebung (GLS) vor. Nach einem Überblick über Methodik und Merkmale (Kapitel 2) erläutern wir die Möglichkeiten des Datenzugangs (Kapitel 3), wobei der Schwerpunkt der Darstellung auf dem kürzlich erschienenen Scientific-Use-File der Erhebung 2001 liegt. Dabei verdeutlichen wir, dass die Anonymisierung von LEED für die Datenproduzenten eine besondere Herausforderung darstellt, da die Balance zwischen Schutz der Daten und Erhalt des Analysepotenzials schwieriger erreichbar ist als bei Daten, die nur Informationen entweder über Unternehmen oder über Beschäftigte enthalten. Kapitel 4 fasst die Ergebnisse einiger mit der GLS durchgeführter Untersuchungen zusammen und zeigt damit das in den Daten enthaltene Forschungspotenzial auf. Abschließend geben wir einen kurzen Ausblick auf die neuen Perspektiven, die sich durch die geplante Verknüpfung der GLS mit anderen Wirtschaftsstatistiken ergeben, wobei hier allerdings noch einige technische Fragen zu klären sind.

---

<sup>1</sup> Forschungsdatenzentrum der Statistischen Ämter der Länder, hhafner@statistik-hessen.de

<sup>2</sup> Fachhochschule Mainz, rainer.lenz@fh-mainz.de

## 2. Methodik und Merkmale

Die GLS wird von den Statistischen Ämtern des Bundes und der Länder seit 1951 durchgeführt. Nachdem die Erhebung in der Vergangenheit in unregelmäßigen Abständen mit einer großen Lücke zwischen 1978 und 1990 stattfand, ist die Periodizität zukünftig vierjährlich. Auf Grundlage einer Verordnung der EG von 1999 findet die Erhebung in allen EU-Ländern statt. Somit liegen europaweit vergleichbare Daten vor. Da die meisten Länder die letzte Erhebung für das Jahr 2002 durchgeführt haben, findet die nächste Erhebung für das Jahr 2006 statt. Dafür werden Anfang 2007 die Erhebungsbögen den Betrieben zugesandt. Die Aufbereitung der Daten wird voraussichtlich in der zweiten Jahreshälfte 2008 abgeschlossen sein.

Die Angaben zu Arbeitszeit und Verdienst beziehen sich immer auf den Monatsmonat Oktober. Für Deutschland lieferten 2001 gut 22.000 Betriebe Angaben zu über 846.000 Beschäftigten. Zum Berichtsbereich gehören Betriebe des Produzierenden Gewerbes und ausgewählte Teile des Dienstleistungsbereiches. Neben Handel und Kredit- und Versicherungsgewerbe gibt es seit 2001 Daten für Gastgewerbe, Verkehr und Nachrichtenübermittlung, Grundstücks- und Wohnungswesen, Vermietung beweglicher Sachen und Erbringung von Dienstleistungen überwiegend für Unternehmen. Auch bei den Beschäftigten erfolgte eine Ausweitung gegenüber früheren Erhebungen: Auszubildende, geringfügig Beschäftigte und Arbeitnehmer in Altersteilzeit sind erstmals 2001 erfasst worden.

Die GLS ist eine zweistufige Stichprobenerhebung. In der ersten Auswahlstufe wird aus den Betrieben eine geschichtete Zufallsstichprobe gezogen. Die Schichtung erfolgt dabei nach 17 Regionen (Bundesländer, Berlin getrennt nach West und Ost), 64 Wirtschaftsgruppen und 7 Beschäftigtengrößenklassen. In der zweiten Stufe werden in den ausgewählten Betrieben die einzubeziehenden Beschäftigten über die Personalnummer aus den Personallisten ermittelt. Dazu wird von den Statistischen Ämtern eine Startzahl und ein Auswahlabstand vorgegeben. Die Verteilung des Stichprobenumfanges auf die Bundesländer erfolgt derart, dass die Ergebnisse für jede Region einen vergleichbaren Standardfehler besitzen. Demnach ist der Auswahlatz für kleinere Bundesländer höher als für größere. Für eine tiefere regionale Untergliederung als nach Bundesländern sind keine repräsentativen Aussagen möglich. Da im Datensatz aber auch die Gemeindekennziffer des Betriebssitzes enthalten ist, könnten darüber z. B. zusätzliche Informationen auf Kreisebene angespielt werden wie etwa die Siedlungsstrukturtypen des Bundesamtes für Bauwesen und Raumordnung (BBR) oder regionale Arbeitslosenquoten.

Es gibt getrennte Erhebungsbögen für die Angaben zum Betrieb und jeweils einen Bogen (oder bei größeren Betrieben mehrere) für die Arbeiter und Angestellten. Für einen Betrieb bzw. das Unternehmen, zu dem der Betrieb gehört, liegen folgende Angaben vor:

- Wirtschaftszweig nach WZ 93 (deutsche Systematik, von NACE Rev. 1 abgeleitet)

- Einfluss der öffentlichen Hand auf die Unternehmensführung: Kein Einfluss, eingeschränkter Einfluss bis 50 %, beherrschender Einfluss über 50 %
- Anzahl tätige Personen des Unternehmens
- Anzahl Beschäftigte des Betriebes differenziert nach Arbeitern und Angestellten und Geschlecht
- Schlüssel der Tarifverträge, die für den Betrieb gelten: Daraus kann man u. a. ablesen, ob für den Betrieb ein Kollektiv- oder Firmentarifvertrag oder eine Betriebsvereinbarung gilt
- Hochrechnungsfaktoren: Diese Faktoren dienen zur freien Hochrechnung der Daten auf die Grundgesamtheit der Betriebe bzw. der Arbeitnehmer. Sie ergeben sich für die Betriebe als Quotient der Anzahl der Unternehmen in der Schicht in der Grundgesamtheit und der Anzahl der verwertbaren Antworten in der Schicht, für die Arbeitnehmer als Quotient aus der Anzahl der Beschäftigten des Betriebes insgesamt und der Anzahl der in die Erhebung einbezogenen Beschäftigten.

Für die Beschäftigten werden folgende Angaben erfasst:

- tarifliche Gehalts- oder Lohngruppe, sofern nach Tarifvertrag bezahlt wird
- Leistungsgruppe: Bei Bezahlung nach Tarifvertrag werden die einzelnen Gehaltsklassen den Leistungsgruppen zugeordnet. Erfolgt die Bezahlung nach freier Vereinbarung, so muss der Betrieb anhand der ausgeübten Tätigkeit entscheiden, zu welcher Leistungsgruppe der Beschäftigte gehört. Bei den Angestellten erfolgt eine Differenzierung nach 6 Gruppen, bei den Arbeitern nach 4 Gruppen.
- soziodemographische Merkmale: Geschlecht, Geburtsmonat und -jahr
- Angaben zu Tätigkeit und Qualifikation: Der fünfstellige Schlüssel aus dem Versicherungsnachweis der Sozialversicherung wird in den Bogen eingetragen. Die ersten drei Stellen bezeichnen die ausgeübte Tätigkeit, die vierte die Stellung im Beruf und die letzte die Ausbildung.
- Arbeitszeit: Vertraglich vereinbarte wöchentliche Arbeitszeit, Mehrarbeitsstunden
- Verdienst: Bruttomonatsverdienst, Verdienst für Mehrarbeit, Zulagen für Schicht-, Nacht- und Sonntags-/Feiertagsarbeit, Bruttojahresverdienst, jährliche Sonderzahlungen
- Urlaubsanspruch in Tagen

Einige zusätzliche Merkmale werden auf Grundlage eines nationalen Gesetzes nur für das Produzierende Gewerbe, Handel und Kredit- und Versicherungsgewerbe erhoben. Dazu gehören Lohnsteuer und Sozialabgaben, Nettoverdienst sowie Steuerklasse und die Anzahl der Kinderfreibeträge.

Weitere Informationen zu Methodik und Merkmalen der GLS 2001 findet man in Frank-Bosch 2003 und in den Metadaten im Internetangebot der Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder:

<http://www.forschungsdatenzentrum.de/bestand/gls/suf/2001/index.asp>

### 3. Möglichkeiten des Datenzugangs

In den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder stehen die Einzeldaten der GLS 2001 und GLS 1995 für ganz Deutschland zur On-Site-Nutzung an den Gastwissenschaftlerarbeitsplätzen an allen Standorten und mittels kontrollierter Datenfernverarbeitung zur Verfügung. Bei einem Vergleich der aktuellen Daten mit den älteren ist zu berücksichtigen, dass 1990 komplett nach der alten Wirtschaftszweigklassifikation für das Verarbeitende Gewerbe (SYPRO) erhoben wurde; 1995 ist für die Mehrzahl der Betriebe bereits der Code nach WZ93 vorhanden, aber nicht für alle. Für den Umstieg von SYPRO auf WZ93 gibt es zwar Umschlüsselungsbrücken, aber die Zuordnung von den alten zu den neuen Wirtschaftszweigen ist in vielen Fällen nicht eindeutig möglich.

Seit Anfang 2007 ist ein Scientific-Use-File der Erhebung 2001 erhältlich, der in einem Kooperationsprojekt zwischen dem Hessischen Statistischen Landesamt und dem Statistischen Bundesamt entstanden ist. Damit das Ergebnis für einen möglichst großen Nutzerkreis von Interesse ist, wirkte ein wissenschaftlicher Begleitkreis bei der Konzeption der Datei beratend mit.

Zunächst war die entscheidende Frage, welche regionalen Einheiten ausgewiesen werden. Aufgrund des Stichprobendesigns ist eine Analyse der Daten unterhalb der Ebene der Bundesländer nicht möglich. Für kleinere Bundesländer gibt es darüber hinaus Geheimhaltungsprobleme, wenn man eine aussagekräftige Gliederung der Wirtschaftszweige erhalten möchte. Getestet wurden zwei Alternativen mit fünf bzw. acht jeweils aus Zusammenfassungen benachbarter Bundesländer bestehender Regionen. Je nach Modell wurden etwa 30 – 40 Wirtschaftsgruppen ausgewiesen. Ferner wurde für die besonders gefährdeten großen Unternehmen mit 1000 und mehr Beschäftigten und die jeweils drei größten Unternehmen pro Region und Wirtschaftsgruppe eine Mikroaggregation der Beschäftigtenanzahlen durchgeführt, wobei jede Gruppe aus mindestens drei Unternehmen bestand.

Überschneidungsmerkmale mit kommerziellen Datenbanken wie etwa der Markus-Datenbank sind die Regionalangabe, der Wirtschaftszweig, die Anzahl der Beschäftigten des Unternehmens und in eingeschränkter Form die Beteiligung der öffentlichen Hand am Unternehmen. Das letztgenannte Merkmal lässt sich in der Markus-Datenbank mit Gesellschaftertyp – Behörden, Staaten, Regierungen vergleichen. Das Reidentifikationsrisiko kann mittels geeigneter Matchingexperimente zwischen externen Daten und Zieldaten ermittelt werden. Das Ziel des Datenangreifers besteht hier darin zu entscheiden, ob ein Paar zweier Objekte aus den beiden Datenquellen zu demselben Betrieb bzw. Unternehmen gehört. Umgangssprachlich kann der Begriff des Matchings verstanden werden als die Zusammenführung von Informationen zweier Objekte verschiedener Datenquellen. Zu diesem Zweck ist ein vernünftiges Ähnlichkeitsmaß erforderlich. Technische Details zu diesem Ansatz finden sich in Lenz (2006).

Im vorliegenden Falle stellte sich heraus, dass einige Wirtschaftsbereiche bei einer Ausweisung von acht Regionen zu stark gefährdet wären. Somit haben wir uns für fünf Regionen entschieden und für eine Herabsetzung der unteren Schwelle, ab der die Beschäftigtenangabe mikroaggregiert wurde. Das Merkmal „Einfluss der öffentlichen Hand“ war unkritisch in Bezug auf das Reidentifikationsrisiko und konnte somit im Original belassen werden.

Das Reidentifikationsrisiko für einzelne Arbeitnehmer ist vernachlässigbar, da es hier kein systematisch aufbereitetes Zusatzwissen gibt. Zudem hat ein potenzieller Datenangreifer keine Kenntnis über die Teilnahme einer Person an der Erhebung. Allerdings lassen sich aufgrund der ausgeübten Tätigkeit einer Person Rückschlüsse auf den Wirtschaftsbereich des zugehörigen Betriebes ziehen, sodass die Anonymisierung der Betriebe teilweise wieder aufgehoben wird.

Wir erläutern die mögliche Strategie eines Datenangreifers im Folgenden am Beispiel der stark korrelierten Merkmale „Wirtschaftsgruppe“ und „ausgeübte Tätigkeit“. Eine technische Abhandlung findet sich in Hafner und Lenz 2006.

- (1) Für jede in der anonymisierten Datei zusammengefasste Wirtschaftsgruppe kennt der Datenangreifer die einzelnen Wirtschaftsgruppen, aus denen sie sich zusammensetzt.
- (2) Für jeden der einzelnen Wirtschaftszweige kann sich der Datenangreifer die Verteilung der ausgeübten Tätigkeiten beschaffen, z. B. aus einer Veröffentlichung der statistischen Ämter oder mittels kontrollierter Datenfernverarbeitung.
- (3) Der Datenangreifer vergleicht für jeden Betrieb die Verteilung der ausgeübten Tätigkeiten in der Erhebung mit den Verteilungen der potenziell möglichen Wirtschaftszweige und ordnet den Betrieb dem Wirtschaftszweig zu, für den die Abweichungen am geringsten sind.

Kann der Datenangreifer mit dieser Strategie viele Betriebe richtig zuordnen, so ergibt sich ein hohes Reidentifikationsrisiko für die Betriebe der betroffenen Wirtschaftszweige.

Die Konsequenz für den Datenanbieter ist in einem solchen Falle, dass er vor der Bereitstellung der Daten Berufe, die eine Zuordnung zu einem Wirtschaftszweig unangemessen leicht ermöglichen, mit anderen Berufen zusammenfassen muss.

Im Scientific-Use-File der GLS werden gut 60 Berufsgruppen ausgewiesen. Zu den einzelnen Wirtschafts- und Berufsgruppen sowie weiteren Anonymisierungsmaßnahmen siehe Hafner, Lenz und Mischler 2007. Der Scientific-Use-File ist über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder für die Wissenschaft erhältlich. Die hierzu notwendigen Anträge stehen als Download im Internet unter [www.forschungsdatenzentren.de](http://www.forschungsdatenzentren.de) zur Verfügung.



#### **4. Analysen mit der Gehalts- und Lohnstrukturerhebung**

Die bisherigen Analysen mit der GLS lassen sich zu einem großen Teil den beiden Schwerpunkten „Geschlechtsspezifische Verdienstunterschiede“ und „Verdiensthöhe in Abhängigkeit von Tarifbindung“ zuordnen. Sie entstanden einerseits im Rahmen eines Kooperationsprojektes der Universität Hannover, des Niedersächsischen Landesamtes für Statistik und des Niedersächsischen Ministeriums für Wirtschaft, Technologie und Verkehr, zum anderen innerhalb eines Pilotprojektes des Statistischen Bundesamtes zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und amtlicher Statistik. Außerdem gab es 2004 eine gemeinschaftliche Analyse des Statistischen Landesamtes Baden-Württemberg mit dem Institut für Angewandte Wirtschaftsforschung Tübingen. Erste Ergebnisse von Projekten im Rahmen des Forschungsdatenzentrums der Statistischen Ämter der Länder liegen ebenfalls bereits vor.

##### **4.1 Geschlechtsspezifische Verdienstunterschiede**

Stephan 1997 untersucht mit den niedersächsischen Daten der GLS 1990, inwieweit die Lohnunterschiede zwischen Männern und Frauen darauf zurückzuführen sind, dass einerseits die Geschlechter einen unterschiedlichen Zugang zu Leistungsgruppen, Betrieben und Arbeitsplatzbereichen haben und dass andererseits das Einkommen bei steigendem Frauenanteil in einem Betrieb oder in einer Leistungsgruppe zurückgeht. Sie verwendet dabei ein von Groshen 1991 entwickeltes Zerlegungsverfahren. Bei diesem wird zunächst mit der Methode der kleinsten Quadrate eine Einkommensfunktion geschätzt, bei der neben einer Dummyvariablen für das Geschlecht die Frauenanteile in der Leistungsgruppe des Beschäftigten, im Betrieb und im Arbeitsplatzbereich als erklärende Variablen aufgenommen werden. Die Differenz der geschätzten Durchschnittseinkommen von Männern und Frauen wird dann zerlegt in einen Teil, der nur vom Geschlecht abhängig ist und drei Komponenten, die von den Frauenanteilen in den Bereichen abhängen.

Es zeigt sich, dass bei steigendem Frauenanteil in den einzelnen Bereichen die Einkommen immer sinken. Da sich Frauen öfter für Arbeitsbereiche mit einem bereits hohen Frauenanteil entscheiden als Männer, lassen sich dadurch die Lohnunterschiede erklären. Ein gutes Drittel des Lohndifferentials lässt sich auf den Frauenanteil in der Leistungsgruppe zurückführen, etwa ein Fünftel auf den Anteil im Arbeitsplatzbereich, während der Anteil im Betrieb mit 5 % nur eine geringere Rolle spielt. Das verbleibende gute Drittel kann mit diesem Modell nicht erklärt werden. Von Kulmiz 2001 untersucht geschlechtsspezifische Lohnunterschiede ebenfalls mit den Daten der GLS 1990, allerdings für das ganze frühere Bundesgebiet. Sie ermittelt mit Hilfe einer für Frauen und Männer getrennt durchgeführten schrittweisen Regression Einkommensfunktionen. Dabei verwendet sie als erklärende Variablen u. a. das Alter, das quadrierte Alter, die Unternehmenszugehörigkeit (linear und quadriert), die Leistungsgruppe, die Ausbildung, 33 Berufsklassen, 25 Wirtschaftszweige, die Arbeitszeit, den aus der Steuerklasse abgeleiteten Familienstand so-

wie Interaktionsterme aus, zum einen Alter und Unternehmenszugehörigkeit und zum anderen Ausbildung und Leistungsgruppe. Die Differenz der durchschnittlichen geschätzten Einkommen für Frauen und Männer wird dann nach dem Verfahren von Oaxaca und Blinder (Oaxaca 1973, Blinder 1973) in einen Ausstattungs- und einen Diskriminierungsanteil zerlegt. Bei den Arbeitern lassen sich nach der Zerlegung 41 % der Differenz erklären, davon knapp 23 % durch die Leistungsgruppe, 13 % durch den Wirtschaftszweig und 7 % durch die Ausbildung. Für die Angestellten lassen sich sogar 64 % durch die Ausstattung der Beschäftigten bzw. die verwendeten Merkmale des Betriebes erklären. Auch hier geht der größte Teil mit über 35 % auf das Konto der Leistungsgruppe, durch Alter und Wirtschaftszweig lassen sich jeweils 8 % erklären.

#### **4.2 Der Einfluss von Tarifverträgen auf Verdienstunterschiede**

Stephan und Gerlach 2005 untersuchen mit den Erhebungen 1990, 1995 und 2001 für Niedersachsen Verdienstunterschiede im Verarbeitenden Gewerbe in Unternehmen, die nach Kollektivtarifverträgen, Firmentarifverträgen und nach freier Vereinbarung bezahlen. Sie verwenden ein Mehrebenenmodell, das für die Beschäftigten die Dauer der Unternehmenszugehörigkeit und die aus Ausbildung und Alter errechnete potenzielle Berufserfahrung einbezieht; auf Firmenebene werden außer den Angaben zu den Tarifverträgen der Frauenanteil, die Unternehmensgröße und die durchschnittliche Schulbildung berücksichtigt. Analysiert wird, von welchen Faktoren der Basislohn (Lohn, den der durchschnittliche Beschäftigte im Unternehmen erhält) abhängt. Der Basislohn ist in Firmen, die einen Tarifvertrag anwenden, höher, 1990 etwa 3 bis 4 %, 2001 gar 11 bis 12 %. Dabei ist die Differenz bei Verwendung von Kollektivtarifverträgen noch leicht größer als bei Firmentarifverträgen. Von 1990 bis 2001 haben sich aber die durch Qualifikation und Unternehmenszugehörigkeit bedingten Unterschiede verringert. Die Wissenschaftler begründen dies damit, dass bei einer schlechteren wirtschaftlichen Lage, wie sie 2001 im Vergleich zu 1990 bestand, zusätzliche Qualifikationen weniger stark vergütet werden. Weitere Ergebnisse sind, dass der Basislohn bei steigendem Frauenanteil sinkt und bei höherer durchschnittlicher Bildung steigt. Das letztere hängt wahrscheinlich damit zusammen, dass gut zahlende Unternehmen Hochqualifizierte besonders stark anziehen. Der Lohnabstand von Frauen zu Männern ist in Unternehmen mit einem Tarifvertrag ca. 5 % geringer, von 1990 bis 2001 reduzierte er sich in Unternehmen, die nach freier Vereinbarung vergüten, von 21 auf 17 %. In Gerlach und Stephan 2006 untersuchen sie für die gleichen Unternehmensgruppen die Lohnspreizung bei Arbeitern. Wie erwartet ist diese am größten in Firmen, die nach freier Vereinbarung vergüten, da die Gewerkschaften bei der Aushandlung von Tarifverträgen darum bemüht sind, eine möglichst komprimierte Verdienststruktur zu erreichen. Bei Anwendung von Kollektivtarifverträgen betrug die durchschnittliche Lohnspreizung 19 % in 1990 und 20 % in 1995 und 2001, während sie bei freier Vereinbarung 22 % in 1990, 25 % in 1995 und 24 % in 2001 erreichte. Zerlegt man die Varianz der Löhne in Firmeneffekte und in von individuellen Charakteristika der Arbeiter abhängige Anteile, so überwiegen die

Anteile der Firmeneffekte bei der Anwendung von Kollektivtarifverträgen, während die individuellen Leistungsmerkmale bei Bezahlung nach freier Vereinbarung den größten Einfluss haben.

Bechtel, Heinbach und Strotmann 2006 untersuchen die gleiche Fragestellung mit den Daten der GLS 2001 für Baden-Württemberg und kommen dabei zu ähnlichen Ergebnissen. Heinbach 2005 differenziert bei Kollektivtarifverträgen weiter danach, ob sie Öffnungsklauseln enthalten, die z. B. eine Abweichung bezüglich Löhnen, Arbeitszeit oder Sonderzahlungen ermöglichen. Dazu wertete er zunächst Tarifverträge aus, die vom Ministerium für Arbeit und Soziales und vom WSI-Tarifarchiv zur Verfügung gestellt wurden. Die daraus zusammengestellten Informationen wurden mit Hilfe des vom Statistischen Bundesamt zur Verfügung gestellten Tarifleitbandes, welches die in der GLS enthaltenen Tarifvertragsschlüssel und die zugehörigen Bezeichnungen der Verträge enthält, mit den Daten der GLS verknüpft. Eine Unsicherheit bei der Untersuchung besteht darin, dass man keine Informationen darüber hat, welche Firmen von den Öffnungsklauseln Gebrauch machen. Deskriptive Analysen zeigen, dass die Durchschnittsverdienste geringer sind, wenn Öffnungsklauseln im Tarifvertrag enthalten sind. In Varianzanalysen und Regressionsmodellen bestätigt sich aber nur für die Arbeiter ein signifikanter Einfluss der Öffnungsklauseln auf die Lohnverteilung, während die Koeffizienten für die Angestellten nicht signifikant sind. Die Untersuchung soll nun auf ganz Deutschland ausgedehnt werden, wobei auch die Daten von 1995 mit herangezogen werden.

Ebenfalls mit dem Einfluss verschiedener Tarifregime auf die Lohnverteilungen beschäftigen sich die beiden Arbeiten von Fitzenberger, Kohn und Lembcke 2007 bzw. Kohn und Lembcke 2006. Im Papier von 2006 untersuchen sie Verteilungen und Regressionsmodelle auf Individual- und Firmenlevel. Sie betrachten getrennt männliche und weibliche Vollzeitbeschäftigte und weibliche Teilzeitbeschäftigte. Dabei differenzieren sie jeweils nach Arbeitern und Angestellten und nach West- und Ostdeutschland. Im Allgemeinen werden die aufgrund früherer Analysen erwarteten Ergebnisse bestätigt. Es gibt aber Beschäftigtengruppen mit bemerkenswerten Abweichungen. So ist für männliche Angestellte in Vollzeit das durchschnittliche Gehalt bei individueller Vereinbarung am höchsten, während dies für die anderen Beschäftigtengruppen bei Vorliegen eines Firmentarifvertrages der Fall ist. Dieser Effekt wird damit erklärt, dass außertarifliche Angestellte einen ausgeprägten Effekt auf das Gehaltsniveau haben. Für westdeutsche Angestellte ist die Gehaltsspreizung höher bei Firmentarifverträgen als unter Kollektivtarifverträgen, während es im Osten umgekehrt ist. Aber im Allgemeinen sind die diesbezüglichen Unterschiede zwischen West und Ost gering. Die Regressionsmodelle zeigen, dass die Basislöhne bei Vereinbarungen auf Firmenebene deutlich höher sind als bei Kollektivtarifverträgen oder individueller Vereinbarung. Dies deutet darauf hin, dass Gewerkschaften erfolgreicher sind, wenn sie direkt mit dem Arbeitgeber verhandeln. Ferner haben Gewerkschaften in Ostdeutschland einen größeren Einfluss. Dies zeigt sich daran, dass dort die Reduzierung der Lohnspreizung bei Anwendung eines Tarifvertrages deutlich stärker ausfällt als im Westen. Im Aufsatz von 2007 untersuchen sie zusätzlich den Einfluss der Netto-Gewerkschaftsdichte (net uni-

on density, NUD) für Westdeutschland, die als Anteil der beschäftigten Gewerkschaftsmitglieder an allen Beschäftigten eines Arbeitsmarktes definiert ist. Da die Angabe, ob ein Beschäftigter Mitglied einer Gewerkschaft ist, nicht in der GLS enthalten ist, imputieren sie dieses Merkmal mit einem Schätzverfahren, dass Fitzenberger, Kohn und Wang 2006 mit Daten des Sozioökonomischen Panels (SOEP) entwickelt haben. Bei Quantilsregressionen der bedingten Verteilung des logarithmierten Stundenlohns mit individuellen Charakteristika der Beschäftigten, Betriebsmerkmalen und Tarifvertrags-/Gewerkschaftsvariablen als Kovariaten zeigt sich, dass die NUD einen negativen Effekt auf die Lohnhöhe hat. Während beim 10. Perzentil nur ein nichtsignifikanter Effekt von -0,7 % bei einem Anstieg der Gewerkschaftsdichte um 10 % feststellbar ist, beträgt der gleiche Effekt beim 90. Perzentil signifikante -2,4 %.

#### **4.3 Weitere Forschungsarbeiten**

Fitzenberger und Reize 2002 untersuchen für die alten Bundesländer Lohnunterschiede und -wachstum von 1990 bis 1995 differenziert nach verschiedenen Qualifikationsgruppen. Diese Gruppen werden definiert nach den Dimensionen Berufsausbildung (ohne abgeschlossene Berufsausbildung; mit abgeschlossener Berufsausbildung, aber ohne (Fach-) Hochschulabschluss; mit (Fach-) Hochschulabschluss) und Erwerbstyp (Vollzeit erwerbstätige Männer, Vollzeit erwerbstätige Frauen, Teilzeit erwerbstätige Frauen).

Alle Lohnverteilungen sind eingipflig. Erwartungsgemäß liegt die Verteilung für die Vollzeit erwerbstätigen Männer am weitesten rechts und für die Teilzeit erwerbstätigen Frauen am weitesten links. Die Struktur der Verteilungen ändert sich von 1990 bis 1995 kaum. Die Verdienstunterschiede innerhalb der Gruppen nehmen mit steigendem Qualifikationsniveau zu. Besonders ausgeprägt sind diese Unterschiede bei den Teilzeit erwerbstätigen Frauen, während sie sich zwischen Vollzeit tätigen Männern und Frauen nicht wesentlich unterscheiden.

Jirjahn und Stephan 1999 analysieren mit den Daten für Niedersachsen 1990 und 1995 für männliche Beschäftigte, von welchen Faktoren die Höhe der jährlichen Sonderzahlungen abhängt. Mittels Varianzanalysen wird gezeigt, welche Anteile der Differenzen bei den Sonderzahlungen auf den Beschäftigten, den Betrieb, die Branche und den Arbeitsplatzbereich zurückzuführen sind. Die individuellen Merkmale des Beschäftigten wie Ausbildung und Alter erklären nur um die 10 % der Varianz, während insbesondere die Dauer der Betriebszugehörigkeit eine wichtige Rolle spielt. Durch den Betrieb werden bei den Arbeitern 70 bis 90 %, bei den Angestellten 50 bis 60 % der Varianz erklärt; Sektor und Firmengröße stehen in der Erklärungskraft dagegen weit zurück.

Boss, Christensen und Schrader 2005 berechnen aus der GLS 2001 durchschnittliche Verdienste nach Altersgruppen und Qualifikationsstufen, wobei sie nach Geschlecht, alten / neuen Bundesländern und sekundärem und tertiärem Sektor unterscheiden. Die mit Hilfe der Ergebnisse aus den laufenden Verdienstrechnungen und Schätzfaktoren auf der Basis anderer Datenquellen auf 2005 hochgerechneten

Werte verwenden sie für einen Vergleich mit der Höhe des Arbeitslosengeldes II unter der Fragestellung, ob der Anreiz für die Aufnahme einer Beschäftigung hoch genug ist. Insbesondere für gering qualifizierte ALG-II-Bezieher mit Kindern und einem nicht erwerbstätigen Partner wird kein ausreichender Anreiz gesehen, während für Hochqualifizierte und Menschen mit einem erwerbstätigen Partner das Gegenteil zutrifft. Kritisch sehen die Wissenschaftler auch, dass immer mehr Beschäftigte aus dem ersten Arbeitsmarkt verdrängt werden durch „1-Euro-Jobs“ und geringfügige Beschäftigungen. Um den Niedriglohnbereich attraktiver zu machen, wird u. a. eine negative Einkommenssteuer vorgeschlagen, bei der bis zu einem bestimmten Bruttolohn eine Lohnsteuergutschrift bezahlt wird, die mit der fälligen Steuer verrechnet wird.

## **5. Ausblick: GLS-Panel und Integration von Wirtschaftsdaten**

Paneldaten werden zunehmend von der Wissenschaft nachgefragt. Deshalb wird derzeit geprüft, ob die GLS-Erhebungen von 1995 und 2001 miteinander verknüpft werden können. Da die GLS eine Stichprobenerhebung ist, findet aber insbesondere bei den kleineren Betrieben ein ständiger Wechsel der befragten Betriebe statt.

Der neue §13a Bundesstatistikgesetzes erlaubt die Zusammenführung von Daten aus verschiedenen Wirtschaftsstatistiken der amtlichen Statistik. In einem Projekt der Forschungsdatenzentren der amtlichen Statistik und des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) sollen Daten für Unternehmen des Verarbeitenden Gewerbes aus u. a. der Kostenstrukturerhebung, der Investitionserhebung, dem Monatsbericht, den Steuerstatistiken und auch der GLS von 1995 bis 2001 verknüpft werden. In einem zweiten Teilprojekt sollen diese Statistiken mit den Daten über Betriebe und sozialversicherungspflichtig Beschäftigte des IAB verknüpft werden. Da es hierfür noch keine gesetzliche Grundlage gibt, ist zunächst vorgesehen, 20.000 Betriebe / Unternehmen um ihre Zustimmung zu diesem Vorhaben zu bitten.

## **Literaturverzeichnis**

Abowd, John M. und Woodcock, Simon D.: Disclosure Limitation in Longitudinal Linked Data. Doyle, Pat et al. (eds.): Confidentiality, Disclosure, and Data Access, Amsterdam, 215-277.

Alda, Holger, Bender, Stefan und Gartner, Hermann (2005). The linked employer-employee dataset of the IAB (LIAB). IAB Discussion Paper Nr. 06/2005, Nürnberg.

Bechtel, Stephan; Heinbach, Wolf-Dieter und Strotmann, Harald (2006): Tarifbindung, betriebliche Lohnhöhe und Lohnstreuung im Produzierenden Gewerbe Baden-Württembergs, In: Amtliche Mikrodaten für die Sozial- und Wirtschaftswissenschaften, Beiträge zu den Nutzerkonferenzen des FDZ der Statistischen Landesämter 2005, Statistische Landesämter, 159-177.

- Blinder, Alan S. (1973): Wage Discrimination: Reduced Form And Structural Estimates, *The Journal Of Human Resources* VIII, 436-455.
- Boss, Alfred; Christensen, Björn und Schrader, Klaus (2005): Anreizprobleme bei Hartz IV: Lieber ALG II statt Arbeit, *Kieler Diskussionsbeiträge* 421, Institut für Weltwirtschaft, Kiel.
- Fitzenberger, Bernd; Kohn, Karsten und Lembcke, Alexander C. (2007): Union Wage Effects in Germany: Union Density or Collective Bargaining Coverage?, unveröffentlichtes Manuskript.
- Fitzenberger, Bernd; Kohn, Karsten und Wang, Qingwei (2006): The Erosion of Union Membership in Germany: Determinants, Densities, Decompositions, Discussion Paper 2193, IZA.
- Fitzenberger, Bernd und Reize, Frank (2002): Verteilung, Differentiale und Wachstum – Eine Verdienstanalyse für Westdeutschland auf Basis der Gehalts- und Lohnstrukturerhebung, Discussion Paper No. 02-71, ZEW Mannheim, <ftp://ftp.zew.de/pub/zew-docs/dp/dp0271.pdf>.
- Frank-Bosch, Birgit (2003): Verdienststrukturen in Deutschland: Methode und Ergebnisse der Gehalts- und Lohnstrukturerhebung 2001, *Wirtschaft und Statistik* 12 (2003), 1137-1151.
- Gerlach, Knut und Stephan, Gesine (2006): Bargaining Regimes and Wage Dispersion, *Jahrbücher für Nationalökonomie und Statistik* 226/6, 629-645.
- Groshen, Erica L. (1991): The Structure Of The Female/Male Wage Differential, *The Journal Of Human Resources* 26, 455-472.
- Hafner, Hans-Peter und Lenz, Rainer (2006): Anonymisation of Linked Employer Employee Datasets. Theoretical Thoughts and an Application to the German Structure of Earnings Survey, *Proceedings of the Privacy in Statistical Databases conference (PSD'2006)*, december 13-15, Rome.
- Hafner, Hans-Peter; Lenz, Rainer und Mischler, Frauke (2007): Einzeldaten der Gehalts- und Lohnstrukturerhebung 2001 als Scientific-Use-File, *Wirtschaft und Statistik* 2 (2007).
- Heinbach, Wolf-Dieter (2005): Impact of Opening Clauses on Bargained Wages, *IAW-Diskussionspapiere* 22, Tübingen.
- Jirjahn, Uwe und Stephan, Gesine (1999): Betriebliche Sonderzahlungen – Theoretische Überlegungen und empirische Befunde. In: B. Frick, R. Neubäumer & W. Sesselmeier (Hrsg.), *Die Anreizwirkungen betrieblicher Zusatzleistungen*, München und Mering::Rainer Hampp Verlag, S. 33-67.
- Kohn, Karsten und Lembcke, Alexander C. (2006): Wage Distributions by Bargaining Regime – Linked Employer-Employee Data Evidence for Germany, unpublished manuscript.

- Lenz, Rainer (2006): Measuring the disclosure protection of micro aggregated business microdata. Analysis taking the example of German Structure of Costs Survey. Appears in: Journal of Official Statistics 22 (4), Sweden.
- Oaxaca, Ronald (1973): Male-Female Wage Differentials In Urban Labor Markets, International Economic Review 14, 693-709.
- Stephan, Gesine (1997): Eine empirische Analyse der Lohnunterschiede zwischen Frauen und Männern, Statistische Monatshefte Niedersachsen 1 (1997), 5-11.
- Stephan, Gesine und Gerlach, Knut (2005): Wage Settlements and Wage Setting: Evidence from a Multi-Level Model, Applied Economics 37, 2297-2306.
- Von Kulmiz, Leontine (2001): Lohndiskriminierung von Frauen. Eine Analyse mit der Gehalts- und Lohnstrukturserhebung 1990, Wirtschaft und Statistik 5 (2001), 406-415.