

Erste  
Scientific-Use-Files  
aus den Wirtschafts-  
statistiken

 **STATISTISCHE ÄMTER  
DES BUNDES UND DER LÄNDER**

 Institut für  
Angewandte  
Wirtschaftsforschung

Auszüge aus Wirtschaft  
und Statistik 2/05 und 3/05

Statist. Bundesamt - Bibliothek



05-00347

(05-00343)

## Auszüge aus Wirtschaft und Statistik

Ein Scientific-Use-File der Kostenstrukturerhebung im Verarbeitenden Gewerbe

(Wirtschaft und Statistik 2/2005, S. 91 ff.)

Ein Scientific-Use-File der Einzelhandelsstatistik 1999

(Wirtschaft und Statistik 3/2005, S. 197 ff.)

Ein Scientific-Use-File der Umsatzsteuerstatistik 2000

(Wirtschaft und Statistik 3/2005, S. 201 ff.)

Herausgeber: Statistisches Bundesamt, Wiesbaden

Schriftleitung: Johann Hahlen  
Präsident des Statistischen Bundesamtes  
Verantwortlich für den Inhalt:  
Brigitte Reimann,  
65180 Wiesbaden

- Telefon: +49 (0) 6 11/75 20 86
- E-Mail: [wirtschaft-und-statistik@destatis.de](mailto:wirtschaft-und-statistik@destatis.de)

Vertriebspartner: SFG Servicecenter Fachverlage  
Part of the Elsevier Group  
Postfach 4343  
72774 Reutlingen  
Telefon: +49 (0) 70 71/93 53 50  
Telefax: +49 (0) 70 71/93 53 35  
E-Mail: [destatis@s-f-g.com](mailto:destatis@s-f-g.com)

Erscheinungsfolge: monatlich



Allgemeine Informationen über das Statistische Bundesamt und sein Datenangebot erhalten Sie:

- im Internet: [www.destatis.de](http://www.destatis.de)

oder bei unserem Informationsservice  
65180 Wiesbaden

- Telefon: +49 (0) 6 11/75 24 05
- Telefax: +49 (0) 6 11/75 33 30
- [www.destatis.de/kontakt](http://www.destatis.de/kontakt)

© Statistisches Bundesamt, Wiesbaden 2005

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.



Dr. Rainer Lenz, Dr. Daniel Vorgrimler (beide Statistisches Bundesamt),  
Dipl.-Volkswirt Martin Rosemann (Institut für Angewandte Wirtschaftsforschung – IAW, Tübingen)

# Ein Scientific-Use-File der Kostenstrukturerhebung im Verarbeitenden Gewerbe

*Das vom Bundesministerium für Bildung und Forschung geförderte Projekt „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ beschäftigt sich mit der Frage, wie eine faktische Anonymisierung von Unternehmens- und Betriebsdaten gelingen kann. Gemeinsam mit der Wissenschaft wollen die statistischen Ämter das so genannte Wissenschaftsprivileg des Bundesstatistikgesetzes (BStatG) umsetzen, nach dem für die Durchführung wissenschaftlicher Vorhaben faktisch anonymisierte Mikrodaten an Hochschulen und sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung übermittelt werden dürfen. Der folgende Beitrag zeigt, wie aus den Daten für kleine und mittlere Unternehmen der Kostenstrukturerhebung im Verarbeitenden Gewerbe eine faktisch anonyme Datei für die Wissenschaft (so genanntes Scientific-Use-File) angeboten werden kann. Neben traditionellen Anonymisierungsmethoden wird hierzu auf eine Variante der Mikroaggregation, einer wesentlichen Verfahrensgruppe der datenverändernden Anonymisierungsverfahren, zurückgegriffen.*

## Vorbemerkung

Im Jahr 1987 wurde mit § 16 Abs. 6 des Bundesstatistikgesetzes<sup>1)</sup> der Wissenschaft ein privilegierter Zugang zu Mikrodaten der amtlichen Statistik eingeräumt. Hiernach ist die Übermittlung von Einzeldaten an die Wissenschaft erlaubt, sofern diese nur mit unverhältnismäßig hohem Aufwand reidentifiziert werden können. „Unverhältnismäßig“ bedeutet hier, dass der Aufwand einer Reidentifikation deren Nut-

zen übersteigt (faktische Anonymität). Dies impliziert, dass die Deanonymisierung von Einzelangaben in einem faktisch anonymen Datensatz nicht mit absoluter Sicherheit ausgeschlossen werden muss, es aber für einen potenziellen Datenangreifer unattraktiv wäre, diese Deanonymisierung zu versuchen.

Während die Bereitstellung von faktisch anonymisierten Personendaten seit längerem bewährte Praxis ist, treten bei wirtschaftsstatistischen Einzeldaten einige Besonderheiten auf, die bisher eine faktische Anonymisierung problematisch erscheinen ließen. Verglichen mit Personenerhebungen liegen bei Unternehmenserhebungen wesentlich kleinere Grundgesamtheiten zugrunde, sodass die Besetzungszahlen einzelner Gruppen häufig kleiner sind. Die Verteilungen der quantitativen Merkmale sind heterogener und es treten mit größerer Wahrscheinlichkeit dominierende Fälle auf. Des Weiteren sind die Stichprobenauswahlsätze bei Unternehmenserhebungen in der Regel höher als bei Personenerhebungen. Publizitätspflichten der Unternehmen einerseits und kommerzielle Datenbanken andererseits bewirken, dass einem Externen, der Einzeldaten ihrem Träger zuordnen will, bei Wirtschaftsdaten besseres Zusatzwissen zur Verfügung steht als bei Personendaten. Schließlich wird der Nutzen aus der Kenntnis von Unternehmens- und Betriebsdaten wesentlich höher eingestuft als der aus der Kenntnis von Daten aus Personen- und Haushaltserhebungen. Ein rational kalkulierender Datenangreifer, der Kosten und Nutzen seines Zuordnungsversuches abwägt, wird deshalb auch höhere Kosten für eine Deanonymisierung

1) Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 16 des Gesetzes vom 21. August 2002 (BGBl. I S. 3322).

akzeptieren.<sup>2)</sup> Hauptaufgabe des Projektes „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ war es daher festzustellen, ob unter diesen strengen Nebenbedingungen der klassische Zielkonflikt zu lösen ist: Sicherstellung der faktischen Anonymität bei gleichzeitig bestmöglichem Erhalt des Potenzials für wissenschaftliche Analysen. Die Arbeiten des Projektes haben erfreulicherweise gezeigt, dass dies möglich ist und mit diesem Beitrag eine erste faktisch anonyme Datei für die Wissenschaft (ein so genanntes Scientific-Use-File), generiert aus den Daten der Kostenstrukturhebung im Verarbeitenden Gewerbe, angeboten werden kann.<sup>3)</sup> Weitere Scientific-Use-Files mit Daten der Umsatzsteuerstatistik und der Einzelhandelsstatistik sind ebenfalls in Vorbereitung.

Zur Lösung des angesprochenen Zielkonfliktes war neben der Berücksichtigung gängiger Methoden der empirischen Forschung ein Einsatz moderner Technologien zur Simulation möglicher Datenangriffsszenarien mit dem Ziel, die Schutzwirkung von Anonymisierungsmaßnahmen angemessen bewerten zu können, nötig. Im Rahmen des EU-Projektes CASC (Computational Aspects of Statistical Confidentiality) hat das Statistische Bundesamt hierzu ein Verfahren zur Zuordnung zwischen verschiedenen Datenbanken entwickelt, über welches unter anderem in dieser Zeitschrift berichtet wurde.<sup>4)</sup> Das Verfahren wurde bereits erfolgreich auf verschiedene Projektdaten angewendet. Dabei wird versucht, über so genannte Massenfischzüge möglichst viele Merkmalsträger (in diesem Falle Unternehmen) zu reidentifizieren. Bei einem Massenfischzug versucht ein Datenangreifer mit Hilfe einer externen Datenbank als Zusatzwissen so viele Merkmalsträger der Zieldaten wie möglich zu reidentifizieren. Im vorliegenden Fall wurden die kommerziell erhältliche MARKUS-Datenbank (Marketinguntersuchungen) und ein relevanter Auszug aus der Umsatzsteuerstatistik als Zusatzwissen für Simulationen verwendet. Beispielhafte Ergebnisse dieser Szenarien wurden bereits in dieser Zeitschrift veröffentlicht.<sup>5)</sup> Grundsätzlich werden vor der Anwendung des Verfahrens kritische Bereiche in den Daten ausfindig gemacht (z. B. ein spezieller Wirtschaftszweig oder Unternehmen einer bestimmten Größenklasse aus den neuen Bundesländern). Das Verfahren wird dann auf diese Bereiche separat angewendet. Die Ergebnisse haben gezeigt, dass in der Regel eine Unterdrückung oder Vergrößerung von Informationen bei den qualitativen Merkmalen wie zum Beispiel die Zusammenfassung von Wirtschaftsabteilungen oder eine Vergrößerung der Regionalangabe beachtlich zur Anonymisierung beiträgt und eine vergleichsweise schwache Modifikation der im Datensatz vorhandenen quantitativen Merkmale ermöglicht. Diese Ergebnisse wurden auch mittels zusätzlich durchgeführter Einzelangriffe bestätigt.<sup>6)</sup> Bei den für das Scientific-Use-File auf die Originaldaten angewendeten Anonymisierungsmaß-

nahmen wurde daher ein großes Gewicht auf die Behandlung der qualitativen Merkmale gelegt.

Im Folgenden werden Datensatz und angewendete Anonymisierungsmaßnahmen beschrieben.

## 1 Datengrundlage

Inhaltlich liefern die Kostenstrukturhebungen im Verarbeitenden Gewerbe die umfassendsten Informationen zu den Unternehmen im Bereich der Statistik im Produzierenden Gewerbe. Sie dienen als Ausgangspunkt für vielfältige Strukturuntersuchungen, nicht nur in Politik und Verwaltung, sondern auch in der Wirtschaft und ihren Verbänden sowie in der Wissenschaft und vielen anderen gesellschaftlichen Gruppierungen. Die Informationen der Kostenstrukturhebungen bilden darüber hinaus eine unentbehrliche Datengrundlage für die Volkswirtschaftlichen Gesamtrechnungen. Hier werden die Ergebnisse vor allem für die Berechnung der Wertschöpfung und ihrer Komponenten nach Wirtschaftsbereichen im Rahmen der Entstehungsrechnung herangezogen; darüber hinaus liefern sie auch wichtige Informationen für die Input-Output-Rechnung. Im Rahmen der Statistiken im Produzierenden Gewerbe bilden die Kostenstrukturstatistiken u. a. eine Grundlage für die Gewichtung von Produktionsindizes. Die Kostenstrukturhebungen enthalten zahlreiche Informationen zu Beschäftigten, Umsätzen und Kostenstruktur der Unternehmen; die einzelnen in der Datengrundlage vorhandenen Variablen sind der Übersicht zu entnehmen.

Die Kostenstrukturhebung im Verarbeitenden Gewerbe des Jahres 1999 erfasste als hochrechnungsfähige Stichprobe etwa 17 000 Unternehmen mit 20 Beschäftigten und mehr. Die Erhebung wird zentral vom Statistischen Bundesamt durch Versand von Fragebogen, die die Unternehmen ausfüllen, durchgeführt. Die in der Stichprobe gewonnenen Ergebnisse werden auf die Gesamtheit der Unternehmen mit 20 Beschäftigten und mehr hochgerechnet.<sup>7)</sup>

In diesem Beitrag wird eine faktische Anonymisierung der Daten der Kostenstrukturhebung, reduziert auf Unternehmen mit 20 bis einschließlich 249 Beschäftigten, vorgestellt. Damit verbleiben etwa 13 200 Unternehmen in der Datei, die etwa 38% der Grundgesamtheit in ihrer Größenklasse darstellen.

## 2 Anonymisierungsmaßnahmen

Wie bereits erwähnt, wurde den Anregungen der Nutzer folgend ein größeres Gewicht auf die Behandlung der qualita-

2) Siehe Brand, R.: „Anonymität von Betriebsdaten“, Beiträge zur Arbeitsmarkt- und Berufsforschung, Band 237, Institut für Arbeitsmarkt- und Berufsforschung der Bundesanstalt für Arbeit, Nürnberg 2000.

3) Aggregate dieser Erhebung werden in der Fachserie 4, „Produzierendes Gewerbe“, Reihe 4.3 „Kostenstrukturhebung der Unternehmen des Verarbeitenden Gewerbes sowie des Bergbaus und der Gewinnung von Steinen und Erden“ des Statistischen Bundesamtes veröffentlicht.

4) Siehe Lenz, R./Vorgirmler, D.: „Geheimhaltungsmethoden auf dem Prüfstand – eine Analyse anhand der Umsatzsteuerstatistik“ in WiSta 6/2004, S. 639 ff. Eine theoretische Beschreibung des Verfahrens findet sich in Lenz, R.: „A graph theoretical approach to record linkage“, Monographs of Official Statistics-Research in Official Statistics, 2003, S. 324 ff.

5) Siehe Lenz, R./Sturm, R./Vorgirmler, D.: „Maße für die faktische Anonymität von Mikrodaten“ in WiSta 6/2004, S. 621 ff.

6) Zu durchgeführten Einzelangriffen auf die Kostenstrukturhebung siehe Vorgirmler, D.: „Re-Identifikationsmöglichkeiten am Beispiel eines konkreten Angriffsszenarios“ in Ronning, G./Gross, R. (Hrsg.): „Anonymisierung wirtschaftsstatistischer Einzeldaten“, Band 42 der Schriftenreihe „Forum der Bundesstatistik“, Wiesbaden 2003, S. 40 ff.

7) Für eine detaillierte Beschreibung siehe „Beschreibung der Kostenstrukturhebung im Verarbeitenden Gewerbe“, Zwischenbericht zum Projekt „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“, Wiesbaden, Juni 2002.

In der Kostenstrukturerhebung vorhandene Variablen

1. Wirtschaftsabteilung [WZ 93<sup>1)</sup>] auf Zweistellerebene)
2. Regionalbezug (Ost-West-Klassifizierung)
3. Beschäftigtengrößenklasse  
07 = 20 – 49 Beschäftigte  
08 = 50 – 99 Beschäftigte  
11 = 100 – 249 Beschäftigte  
14 = 250 – 499 Beschäftigte  
17 = 500 – 999 Beschäftigte  
22 = 1 000 und mehr Beschäftigte
4. Teilzeitbeschäftigte
5. Teilzeitbeschäftigte umgerechnet in Vollezeiteinheiten
6. Tätige Personen insgesamt
7. Umsatz aus eigenen Erzeugnissen
8. Umsatz aus Handelsware
9. Gesamtumsatz (entspricht nicht der Summe aus 7. und 8.)
10. Anfangsbestand an unfertigen und fertigen Erzeugnissen aus eigener Produktion gemessen am Umsatz aus eigenen Erzeugnissen
11. Endbestand an unfertigen und fertigen Erzeugnissen aus eigener Produktion gemessen am Umsatz aus eigenen Erzeugnissen
12. Gesamtleistung/Bruttoproduktionswert
13. Anfangsbestand an Rohstoffen und sonstigen fremdbezogenen Vorprodukten, Hilfs- und Betriebsstoffen gemessen am Umsatz aus eigenen Erzeugnissen
14. Endbestand an Rohstoffen und sonstigen fremdbezogenen Vorprodukten, Hilfs- und Betriebsstoffen gemessen am Umsatz aus eigenen Erzeugnissen
15. Verbrauch an Rohstoffen
16. Energieverbrauch
17. Anfangsbestand an Handelsware gemessen am Umsatz aus Handelsware
18. Endbestand an Handelsware gemessen am Umsatz aus Handelsware
19. Einsatz an Handelsware
20. Bruttolohn und -gehaltsumme
21. Gesetzliche Sozialkosten
22. Sonstige Sozialkosten
23. Kosten für Leiharbeitnehmer
24. Kosten für Lohnarbeiten
25. Kosten für Reparaturen
26. Mieten und Pachten
27. Sonstige Kosten
28. Fremdkapitalzinsen
29. Kosten insgesamt
30. Bruttowertschöpfung zu Faktorkosten
31. Nettowertschöpfung zu Faktorkosten
32. Gesamtaufwendungen für innerbetriebliche Forschung und Entwicklung
33. Anzahl der für Forschung und Entwicklung eingesetzten Lohn- und Gehaltsempfänger

1) Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ 93).

tiven Merkmale gelegt, um einer stärkeren Verfremdung der Daten bei den quantitativen Merkmalen vorzubeugen.<sup>8)</sup>

### Traditionelle Anonymisierungsverfahren

In einem ersten Schritt wurde auf das ursprünglich im Merkmalskanon vorhandene Merkmal „Tätige Inhaber“ verzichtet, da es sich im Laufe der Projektarbeiten als besonders reidentifikationsgefährdet und für wissenschaftliche Analysen als wenig wertvoll herausgestellt hat.<sup>9)</sup> Das Merkmal „Angestellte und Arbeiter“ wurde daher als Differenz aus „Tätige Personen insgesamt“ und „Tätige Inhaber“ eben-

falls aus dem ursprünglichen Merkmalskanon herausgenommen. Das Merkmal „Bestandsveränderungen“ wurde entfernt, da es aus inhaltlichen Gründen verzichtbar war.

Besonders geeignet für Reidentifikationen sind regionale Angaben. Der Erhalt solcher Merkmale in einem Scientific-Use-File stellt daher für die Anonymisierung ein schwieriges Unterfangen dar. Zu Beginn des Projektes wurde die Möglichkeit ausgeschlossen, ein Scientific-Use-File zu erstellen, das administrative Gebietsangaben auf der Ebene der Bundesländer oder gar einer tieferen Gliederungsebene enthält. Da aber die Auswertung nach Regionen einen wichtigen Analysebereich darstellt, wurde nach alternativen Möglichkeiten gesucht und dies vor dem Hintergrund, gleichzeitig auf datenverändernde Maßnahmen bei den quantitativen Merkmalen weitestgehend verzichten zu wollen. Als erste Möglichkeit wurde der administrative Gebietsschlüssel durch den nichtadministrativen siedlungsstrukturellen Kreistyp BBR9 und den siedlungsstrukturellen Regionstyp BBR3 ersetzt. Diese Vergrößerung der Regionalinformation führte zu einer sehr deutlichen Verbesserung der Schutzwirkung.<sup>10)</sup> Allerdings sprach sich der zu Projektbeginn eingerichtete Wissenschaftliche Begleitkreis dafür aus, anstelle dieser nicht-administrativen Schlüssel eine Ost-West-Klassifizierung einzuführen. Diese zweite Möglichkeit wurde schließlich in das Scientific-Use-File aufgenommen.

Die Daten der Kostenstrukturerhebung im Verarbeitenden Gewerbe wurden nach der Klassifikation der Wirtschaftszweige, Ausgabe 1993 (WZ 93), auf der Vierstellerebene (Klassen) erhoben und aufbereitet. Diese Klassifikation ist aus der europäischen Wirtschaftszweigklassifikation NACE Rev. 1 abgeleitet, die aufgrund der NACE-Verordnung des Rates der Europäischen Gemeinschaften seit 1995 in allen Mitgliedstaaten der Europäischen Union sowohl für die Erhebung als auch für die Darstellung der statistischen Daten anzuwenden ist.<sup>11)</sup> Das Kodierungssystem der WZ 93 unterscheidet zwischen Abschnitten (Buchstaben A–Q), Unterabschnitten (Buchstaben AA–QA), Abteilungen (Zweisteller), Gruppen (Dreisteller), Klassen (Viersteller) und Unterklassen (Fünfsteller). Der Wirtschaftsbereich „Verarbeitendes Gewerbe sowie Bergbau und Gewinnung von Steinen und Erden“ erstreckt sich über die Abschnitte C und D bzw. – in der numerischen Gliederung – über die Abteilungen 10 bis 37. Im Projekt „Faktische Anonymisierung wirtschaftsstatischer Einzeldaten“ haben sich statistische Ämter und Datennutzer darauf verständigt, bei dem hierarchischen Merkmal WZ 93 die Gliederungstiefe 2 (Zweistellerebene) nicht zu unterschreiten, da hierdurch zum einen eine beachtliche Schutzwirkung [beispielhafte Berechnungen finden sich in dem Beitrag Lenz/Sturm/Vorgrimler<sup>12)</sup>] und zum anderen nach Einschätzung der beteiligten Wissenschaftler für ein Scientific-Use-File eine ausreichende Breite an Analysemöglichkeiten<sup>13)</sup> erhalten wird.

8) Hier ergibt sich eine Parallele zur faktischen Anonymisierung der Daten der Einkommensteuerstatistik, auch in diesem Fall sprach sich der Wissenschaftliche Beraterkreis für einen Erhalt der quantitativen zu Lasten der qualitativen Merkmale aus; siehe Merz, J./Vorgrimler, D./Zwick, M.: „Faktisch anonymisiertes Mikrodatenfile der Lohn- und Einkommensteuerstatistik 1998“ in WiSta 10/2004, S. 1079 ff.

9) Siehe Fußnote 6.

10) Siehe Lenz, R./Vorgrimler, D.: „Matching German Turnover Tax Statistics“, erscheint in der Reihe der Diskussionspapiere des Forschungsdatenzentrums des Statistischen Bundesamtes, Wiesbaden.

11) Für Erhebungen ab dem Jahr 2003 gilt mit dem Branchenschlüssel WZ 2003 eine neue Klassifikation, die Klassifikation der Wirtschaftszweige, Ausgabe 2003.

12) Siehe Fußnote 5.

13) Siehe Rosemann, M./Vorgrimler, D./Lenz, R.: „Erste Ergebnisse faktischer Anonymisierung wirtschaftsstatischer Einzeldaten“ in Allgemeines Statistisches Archiv Band 88/1/2004, S. 73 ff.

In den Veröffentlichungen der statistischen Ämter werden aufgrund von Geheimhaltungsaspekten die Ergebnisse einiger Wirtschaftsabteilungen nur eingeschränkt veröffentlicht. Es handelt sich dabei um Unternehmen der Abteilungen 10, 11, 14, 16, 23, 30, 32, 35 und 37 der WZ 93. Bei den im Projekt durchgeführten Simulationen hat sich bestätigt, dass diese Abteilungen neben den Abteilungen 15, 17, 18, 19, 22 und 34 größerer Geheimhaltung bedürfen. Um diese kritischen Abteilungen im Scientific-Use-File belassen und weitgehend auf datenverändernde Verfahren bei den quantitativen Merkmalen verzichten zu können, werden die Abteilungen 10 (Kohlenbergbau, Torfgewinnung), 11 (Gewinnung von Erdöl und Erdgas, Erbringung damit verbundener Dienstleistungen) und 14 (Gewinnung von Steinen und Erden, sonstiger Bergbau) zum Abschnitt C, die Abteilungen 15 (Ernährungsgewerbe) und 16 (Tabakverarbeitung) zum Unterabschnitt DA, die Abteilungen 17 (Textilgewerbe) und 18 (Bekleidungs-gewerbe) zum Unterabschnitt DB, die Abteilungen 21 (Papiergewerbe) und 22 (Verlags- und Druckgewerbe, Vervielfältigung von bespielten Ton-, Bild- und Datenträgern) zum Unterabschnitt DE, die Abteilungen 30 (Herstellung von Büromaschinen, Datenverarbeitungsgeräten und -einrichtungen) und 31 (Herstellung von

Geräten der Elektrizitätserzeugung, -verteilung u.Ä.) zum Unterabschnitt DL sowie die Abteilungen 34 (Herstellung von Kraftwagen und Kraftwagenteilen) und 35 (Sonstiger Fahrzeugbau) zum Unterabschnitt DM zusammengefasst. Bei den Abteilungen 19 (Leder-gewerbe) und 23 (Kokerei, Mineralölverarbeitung, Herstellung und Verarbeitung von Spalt- und Brutstoffen) wurde das Merkmal WZ 93 unterdrückt. Außerdem wurde die Abteilung 37 (Recycling) aus inhaltlichen und aus Geheimhaltungsgründen herausgenommen. Obwohl die Abteilungen 32 (Rundfunk-, Fernseh- und Nachrichtentechnik) und 33 (Medizin-, Mess-, Steuer- und Regelungstechnik, Optik) ebenfalls zum Unterabschnitt DL zu zählen sind, werden sie im Datensatz separat aufgeführt, da hier die Weitergabe der Zweisteller aus Sicht des Datenschutzes unbedenklich ist. Eine zusammenfassende Aufstellung der im Datensatz vorhandenen Ausprägungen des Merkmals WZ 93 enthält Tabelle 1.

Eindimensionale Mikroaggregation

Die im Datensatz verbleibenden 30 quantitativen Merkmale wurden eindimensional für jedes Merkmal separat mikroaggregiert.<sup>14)</sup> Bei dieser Variante der Mikroaggrega-

Tabelle 1: Verteilung von Gesamtumsatz und Beschäftigten auf die Wirtschaftsabteilungen  
Kostenstrukturerhebung des Verarbeitenden Gewerbes 1999

Wirtschaftsgliederung (H. v.: Herstellung von)	WZ 93 <sup>1)</sup> -Angabe	Unternehmen		Beschäftigte <sup>2)</sup>		Gesamtumsatz <sup>3)</sup>	
		Anzahl	%	Anzahl	%	1 000 DM	
Bergbau und Gewinnung von Steinen und Erden .....	C (10, 11 und 14)	182	1,02	10804	1,17	3 683 797	
Ernährungsgewerbe und Tabakverarbeitung .....	DA (15 und 16)	1704	12,90	137 284	20,70	65 056 450	
Textil- und Bekleidungs-gewerbe ..	DB (17 und 18)	955	6,91	73 529	5,91	18 585 363	
Holzgewerbe (ohne H. v. Möbeln)	20	448	2,81	29 925	2,61	8 206 696	
Papier-, Verlags- und Druckgewerbe .....	DE (21 und 22)	1063	8,49	90 311	8,41	26 421 392	
Chemische Industrie .....	DG (bzw. 24)	628	5,06	53 873	7,84	24 640 473	
H. v. Gummi- und Kunststoffwaren .....	DH (bzw. 25)	659	5,12	54 525	4,24	13 318 912	
Glasgewerbe, Keramik, Verarbeitung von Steinen und Erden ...	DI (bzw. 26)	783	5,47	58 188	5,45	17 118 619	
Metallerzeugung und -bearbeitung .....	27	423	3,73	39 720	4,46	14 004 740	
H. v. Metallerzeugnissen .....	28	1 638	11,94	127 055	9,12	28 663 369	
Maschinenbau .....	DK (bzw. 29)	1 837	14,07	149 732	11,57	36 357 908	
H. v. Büromaschinen, Dv-Geräten und -einrichtungen, Geräte der Elektrizitätserzeugung, -verteilung u. Ä. ....	DL (hier 30 und 31)	757	5,88	62 596	5,05	15 865 903	
Rundfunk-, Fernseh- und Nachrichtentechnik .....	32	219	1,69	17 955	1,39	4 370 507	
Medizin-, Mess-, Steuer- und Regelungstechnik, Optik .....	33	510	3,71	39 513	2,76	8 671 100	
Fahrzeugbau .....	DM (34 und 35)	525	4,24	45 114	3,85	12 101 547	
H. v. Möbeln, Schmuck, Musikinstrumenten, Sportgeräten usw. ....	36	731	5,69	60 516	4,01	12 610 384	
Sonstige .....	19 und 23	164	1,26	13 363	1,46	4 574 646	

1) Klassifikation der Wirtschaftszweige, Ausgabe 1993. – 2) Alle am Jahresende im Betrieb/im Unternehmen tätigen Personen einschließlich tätiger Inhaber/-innen und mithelfender Familienangehöriger (auch unbezahlt mithelfende Familienangehörige, soweit sie mindestens ein Drittel der üblichen Arbeitszeit im Betrieb tätig sind), aber ohne Heimarbeiter/-innen und Zusteller/-innen im Verlagsgewerbe. Einbezogen werden u.a. auch Erkrankte, Urlauber/-innen, Kurzarbeiter/-innen, Streikende, von der Aussperrung Betroffene, Personen in Altersteilzeitregelungen, Auszubildende, Saison- und Aushilfsarbeiter/-innen sowie Teilzeitbeschäftigte. Die Angestellten umfassen auch die kaufmännischen Auszubildenden (einschl. der Auszubildenden in den übrigen nichtgewerblichen Ausbildungsberufen), die Arbeiter/Arbeiterinnen auch die gewerblich Auszubildenden. – 3) Umsatz aus eigener Erzeugung (einschl. Umsatz aus dem Verkauf von Energie und Nebenerzeugnissen und Abfällen sowie Entgelte für industrielle Dienstleistungen, wie Reparaturen, Instandhaltungen, Installationen und Montagen), Umsatz aus Handelsware und sonstigen nichtindustriellen/nichthandwerklichen Tätigkeiten (z.B. Erlöse aus Vermietung und Verpachtung sowie aus Lizenzverträgen, Provisionseinnahmen und Einnahmen aus der Veräußerung von Patenten).

14) Zur Methode der Mikroaggregation und anderen im Projekt untersuchten Methoden siehe Höhne, J.: „Methoden zur Anonymisierung wirtschaftsstatistischer Einzeldaten“ in Ronning, G./Gross, R. (Hrsg.): „Anonymisierung wirtschaftsstatistischer Einzeldaten“, Band 42 der Schriftenreihe „Forum der Bundesstatistik“, Wiesbaden 2003, S. 69 ff.



tion werden zunächst die Merkmalsausprägungen je Merkmal absteigend sortiert. Dann werden tripelweise (aus den Merkmalsausprägungen dreier benachbarter Merkmalsträger) die Durchschnittswerte ermittelt, die Originalwerte durch diese Durchschnittswerte ersetzt und wieder an die ursprüngliche Position zurücksortiert. Falls die Anzahl der Merkmalsträger nicht durch die Zahl 3 teilbar ist, so ist am Ende der absteigend sortierten Liste von Merkmalsausprägungen auch die Bildung einer Gruppe aus vier oder fünf Merkmalsträgern zulässig. Damit ist jede Merkmalsausprägung bei mindestens drei Merkmalsträgern vorhanden. Das hier skizzierte Verfahren ist für das Ziel einer möglichst vielseitigen Datennutzung, sowohl für deskriptive als auch für ökonometrische Auswertungen, das schonendste Verfahren innerhalb der Klasse der Mikroaggregationsverfahren.

Hierzu werden für das vorgestellte Scientific-Use-File zwei Beispiele präsentiert. Zum einen sind in Tabelle 2 die Abweichungen der wichtigsten deskriptiven Statistiken des anonymisierten Datensatzes im Vergleich zum Original dargestellt, wobei sich allenfalls die systematischen, aber geringen Abweichungen der Varianzen als problematisch erweisen können.

Tabelle 2: Veränderungen wichtiger Verteilungscharakteristika durch die Anonymisierung

Mittlerer relativer Fehler		Mittlerer absoluter Fehler	
Arithmetisches Mittel	Varianzen	Korrelationen	Rangkorrelationen
in %		(x 100)	
0,1	4,3	0,8	0,0

Zum anderen werden in Tabelle 3 die Schätzergebnisse für eine Cobb-Douglas-Produktionsfunktion für die kleinen

Tabelle 3: Produktionsfunktion: OLS-Regression

Inputfaktoren	Originalwerte		Anonymisierte Werte	
	Koeffizienten	t-Werte	Koeffizienten	t-Werte
Materialeinsatz .....	0,410	169,76	0,407	168,21
Personalkosten .....	0,351	81,64	0,354	82,20
Externe Dienstleistungen .....	0,056	35,04	0,057	35,31
Sonstige Kosten .....	0,112	43,84	0,114	44,20
Kapitalkosten .....	0,052	18,86	0,049	18,06
Konstante .....	1,775	45,16	1,789	45,27
Anzahl der Beobachtungen .....		12 680		12 685
Bestimmtheitsmaß .....		0,951		0,950
Durchschnittliche relative Abweichung der Koeffizientenwerte .....				2,0%
Anzahl der Koeffizienten außerhalb des 95%-Konfidenzintervalls .....				0
Anzahl der Vorzeichenwechsel .....				0
Anzahl der Koeffizienten, die im Original statistisch signifikant sind, nach der Anonymisierung nicht .....				0
Anzahl der Koeffizienten, die nach der Anonymisierung statistisch signifikant sind, im Original nicht .....				0
Durchschnittliche relative Abweichung der t-Werte .....				1,3%

und mittleren Unternehmen des Verarbeitenden Gewerbes auf Basis Daten der Kostenstrukturerhebung vorgestellt.<sup>15)</sup> Dabei ist erkennbar, dass die Abweichungen auf Basis der mikroaggregierten Daten gegenüber den Originalergebnissen sehr gering sind.

### Schlussbemerkung

Die Wissenschaft genießt seit In-Kraft-Treten des geltenden Bundesstatistikgesetzes von 1987 eine Sonderstellung unter den Datennutzern der Bundesstatistik, das so genannte *Wissenschaftsprivileg*. Es besteht darin, dass für die Durchführung wissenschaftlicher Vorhaben faktisch anonymisierte Mikrodaten vom Statistischen Bundesamt und den Statistischen Ämtern der Länder an Hochschulen oder sonstige Einrichtungen mit der Aufgabe unabhängiger wissenschaftlicher Forschung übermittelt werden dürfen. Der große Vorzug einer solchen Datennutzung für Wissenschaftler liegt darin, dass ihnen die faktisch anonymisierten Einzeldaten übermittelt und diese von ihnen am eigenen Arbeitsplatz ausgewertet werden können.<sup>16)</sup> Ein weiterer Vorteil eines Scientific-Use-Files besteht darin, dass sichergestellt ist, dass verschiedene Nutzer mit demselben Datenmaterial arbeiten können. Dadurch können sich zwischen den Wissenschaftlern Forschungsnetzwerke leichter aufbauen. Darüber hinaus wird das Prinzip der Überprüfbarkeit von wissenschaftlichen Ergebnissen gestärkt.

In Band 42 der Schriftenreihe „Forum der Bundesstatistik“<sup>17)</sup> wurde erstmals aus amtlichen Betriebsdaten ein Public-Use-File<sup>18)</sup>, das einen Ausschnitt von 500 kleinen und mittleren Unternehmen des Verarbeitenden Gewerbes der Kostenstrukturerhebung 1999 beinhaltet, mit eingeschränkten Analysemöglichkeiten der breiten Öffentlichkeit zur Verfügung gestellt. Mit der im vorliegenden Beitrag vorgestellten Datei

15) Zum Vorgehen siehe Rosemann, M.: "Impacts of different versions of micro aggregation on the results of linear estimations", IAW-Workshop "Econometric Analysis of Anonymised Firm Data", Tübingen, März 2004.


16) Zum Vergleich siehe Sturm, R.: „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ in Allgemeines Statistisches Archiv Band 86, 2002, S. 468 ff.

17) CD-ROM als Beilage in Ronning, G./Gnoss, R. (Hrsg.): „Anonymisierung wirtschaftsstatistischer Einzeldaten“, Band 42 der Schriftenreihe „Forum der Bundesstatistik“, Wiesbaden 2003.

18) Im Gegensatz zu einem Scientific-Use-File besteht ein Public-Use-File aus absolut anonymisierten Merkmalsträgern. Diese Daten dürfen nach § 16 Abs. 1 Nr. 4 BStatG jedem Interessierten übermittelt werden. Das Public-Use-File der Kostenstrukturerhebung steht im Internet unter [www.forschungsdatenzentren.de](http://www.forschungsdatenzentren.de) in der Reihe „Campus-File“ als Download kostenlos zur Verfügung.

wird erstmalig eine hochrechnungsfähige Stichprobe über Unternehmen und Betriebe aus Daten der deutschen amtlichen Statistik als standardisiertes Scientific-Use-File für wissenschaftliche Analysen zugänglich gemacht; ein Rohstoff, mit dem die statistischen Ämter arbeiten und der vor allem für empirisch arbeitende Wirtschafts- und Sozialwissenschaftler zur Untersuchung eigener Fragestellungen interessant ist. Die Untersuchungen haben allerdings gezeigt, dass die hier angewendeten Methoden keinen ausreichenden Schutz für Unternehmen mit 250 Beschäftigten und mehr garantieren können. Die erhöhte Anwendung datenverändernder Verfahren, die zu einer stärkeren Verfremdung der einzelnen Merkmalswerte führen (z. B. mehrdimensionale Varianten der Mikroaggregation und entsprechende stochastische Überlagerungen) oder weitere Einschnitte bei den qualitativen Variablen scheinen hier unumgänglich.

Parallel zu den faktisch anonymen Daten der Kostenstrukturerhebung im Verarbeitenden Gewerbe 1999 werden auch Daten der Einzelhandelsstatistik 1999 und der Umsatzsteuerstatistik 2000 angeboten.<sup>19)</sup> Neben der Prüfung der Geheimhaltungsbedingungen wurden die Dateien vom Projektpartner, dem Institut für Angewandte Wirtschaftsforschung (IAW, Tübingen), unterstützt durch den wissenschaftlichen Begleitkreis, auch hinsichtlich des Analysepotenzials intensiv untersucht.

Die Scientific-Use-Files sind für jeweils 65,- Euro 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. 

---

<sup>19)</sup> In den nächsten Ausgaben dieser Zeitschrift werden Beiträge zum Scientific-Use-File der Einzelhandelsstatistik sowie zum faktisch anonymen Mikrodatenfile der Umsatzsteuerstatistik veröffentlicht.

Dipl.-Mathematiker Michael Scheffler

# Ein Scientific-Use-File der Einzelhandelsstatistik 1999

*Die Statistischen Ämter des Bundes und der Länder führen, gefördert vom Bundesministerium für Bildung und Forschung (BMBF), gemeinsam mit der Wissenschaft das Projekt „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ durch. Von der Wissenschaft ist dieses Vorhaben einhellig begrüßt worden. Die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und amtlicher Statistik (KVI) hat die Projektdurchführung in ihrem Abschlussgutachten ausdrücklich empfohlen. Das so genannte Wissenschaftsprivileg des Bundesstatistikgesetzes erlaubt es, faktisch anonymisierte Mikrodatenfiles (so genannte Scientific-Use-Files) an die Wissenschaft weiterzugeben. In diesem Beitrag wird ein im Rahmen des oben genannten Projektes erzeugtes Scientific-Use-File für die Mikrodaten der Einzelhandelsstatistik vorgestellt.*

## Vorbemerkung

Bei der Novellierung des Bundesstatistikgesetzes (BStatG) im Jahr 1987<sup>1)</sup> wurde der Wissenschaft durch die Schaffung des § 16 Abs. 6 ein privilegierter Zugang zu Mikrodaten ermöglicht. Ausgehend vom Wortlaut von § 16 Abs. 6 BStatG gelten Daten als faktisch anonymisiert, „... wenn die Einzelangaben nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft zugeordnet werden können ...“. Demnach ist eine Kosten-Nutzen-Analyse für die Beurteilung der Anonymität maßgeblich.

Im Vergleich zu Personendaten, bei denen die Bereitstellung faktisch anonymisierter Daten längst gängige Praxis ist, gibt es bei Unternehmens- und Betriebsdaten einige Gesichtspunkte, die eine größere Reidentifikationsgefahr begründen: Zunächst sind die Grundgesamtheiten bei Unternehmens- und Betriebserhebungen grundsätzlich kleiner. Weiterhin liegen prinzipiell höhere Stichprobenauswählsätze vor; ab einer bestimmten Umsatzgröße handelt es sich in der Regel um Vollerhebungen. Außerdem existieren mehr einmalige Fälle und es ist mehr Zusatzwissen verfügbar. Diese Punkte führen dazu, dass insbesondere die Anonymisierung größerer Unternehmen schwierig erscheint.

Grundsätzlich gilt es bei der Anwendung von Anonymisierungsverfahren einen „klassischen“ Zielkonflikt zu lösen. Während die statistischen Ämter die Vertraulichkeit der Einzeldaten garantieren müssen, ist die Wissenschaft – im Projekt insbesondere durch das Institut für Angewandte Wirtschaftsforschung (IAW) und einen dort eingerichteten Wissenschaftlichen Begleitkreis vertreten – an Anonymisierungsmaßnahmen interessiert, die das Analysepotenzial der Daten weitestgehend erhalten. Im Laufe des Projektes hat sich herauskristallisiert, dass die Datennutzer Anonymisierungsmaßnahmen, die an den qualitativen Merkmalen ansetzen, gegenüber solchen, die quantitative Merkmale verändern, bevorzugen.

Im ersten Kapitel werden der Datensatz des Scientific-Use-Files der Einzelhandelsstatistik und einige Auswertungs-

1) Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 16 des Gesetzes vom 21. August 2002 (BGBl. I S. 3322).

möglichkeiten vorgestellt. Anschließend werden in Kapitel 2 die Anonymisierungsmaßnahmen erläutert. Kapitel 3 behandelt die Schutzwirkung und das Analysepotenzial des Scientific-Use-Files.

## 1 Datengrundlage

Die Ergebnisse der Jahreserhebung im Einzelhandel liefern wirtschaftspolitisch bedeutsame Informationen über die Struktur, Rentabilität und Produktivität der im Einzelhandel tätigen Unternehmen. Neben der Ermittlung des Rohertrages und der Bruttowertschöpfung sind qualitativ hochwertige Schätzungen für die Vorratsveränderungen in der Wirtschaft möglich. Aus konjunkturpolitischer Sicht können mit den Ergebnissen von Jahreserhebungen aufeinander folgender Jahre sowohl die Beschäftigungssituation als auch die Lohn- und Gehaltsstrukturen beobachtet und analysiert werden. Auf der betriebswirtschaftlichen Ebene lässt sich die Entwicklung von Arbeitsintensität und -produktivität

Merkmale des Scientific-Use-Files der Einzelhandelsstatistik 1999

1. Wirtschaftszweig (WZ 93<sup>1</sup>) auf Drei- bzw. Vierstellerebene
2. Regionalbezug (Ost-West-Klassifizierung)
3. Gesamtumsatz
4. Umsatzanteile in % aus
  - Großhandel
  - Einzelhandel, Reparatur von Gebrauchsgütern
  - Sonstigen Dienstleistungstätigkeiten
  - Herstellung, Verarbeitung, anderen industriellen Tätigkeiten oder aus Land- und Forstwirtschaft, Fischerei
5. Sonstige betriebliche Erträge
6. Einzelhandelsumsatz nach Absatzformen
  - in Verkaufsräumen
  - aus Versandhandel
  - an Verkaufsständen und auf Märkten
  - aus sonstigem Einzelhandel
7. Bestände an Handelsware jeweils am Anfang und Ende des Geschäftsjahres
8. Bestände der Roh-, Hilfs- und Betriebsstoffe jeweils am Anfang und Ende des Geschäftsjahres
9. Bestände der selbst hergestellten und bearbeiteten Halb- und Fertigerzeugnisse jeweils am Anfang und Ende des Geschäftsjahres
10. Bezüge von Handelsware
11. Bezüge von Roh-, Hilfs- und Betriebsstoffen
12. Löhne und Gehälter
13. Sozialabgaben
14. Mieten und Pachten einschließlich Kosten für Operate Leasing
15. Betriebliche Steuern und Abgaben
16. Bezogene Leistungen und andere betriebliche Aufwendungen
17. Bruttoinvestitionen in
  - Grundstücke
  - bestehende Gebäude
  - Errichtung, Umbau und Erweiterung von Gebäuden
  - Maschinen, Einrichtungen und Fahrzeuge
18. Verkäufe von Sachanlagen
19. Wert der im Geschäftsjahr über Finanzierungsleasing erworbenen Sachanlagen
20. Zahl der rechtlich selbstständigen örtlichen Einheiten des Unternehmens am 31. Dezember
21. Zahl der Beschäftigten am 30. September
  - Beschäftigte insgesamt
  - darunter:
  - Lohn- und Gehaltsempfänger
  - Teilzeitbeschäftigte
22. Hochrechnungsfaktor

1) Klassifikation der Wirtschaftszweige, Ausgabe 1993.

2) Das sind die Unternehmen mit einem Mindestjahresumsatz von 250 000 Euro.

überprüfen. Weiterhin sind die Investitionen und ihre Veränderungen wichtiger Indikator für die längerfristige Umsatzerwartung eines Unternehmens. Für die längerfristige Analyse ist es wichtig, die Generierung von faktisch anonymisierten Datenfiles für weitere Jahre in Angriff zu nehmen. Eine Auflistung aller Merkmale des Scientific-Use-Files findet man in der Übersicht.

Die jährliche Einzelhandelsstatistik erfasst für das Jahr 1999 23 510 Unternehmen in der Stichprobe. Diese repräsentieren knapp 300 000 Unternehmen des Einzelhandels mit einem Umsatz von rund 300 Mrd. Euro. Betrachtet man nur die Monatsmelder<sup>2)</sup>, so repräsentieren etwa 14 500 Unternehmen in der Stichprobe über 110 000 Unternehmen der Einzelhandelsstatistik, die einen Umsatz von rund 285 Mrd. Euro tätigen. Aufgrund der relativ hohen Anzahl von Unternehmen mit weniger als 50 Beschäftigten kann für diese eine faktische Anonymisierung nahezu ohne datenverändernde Verfahren erreicht werden. Dagegen wurde für die Unternehmen mit 50 Beschäftigten und mehr bislang noch keine Anonymisierungsmethode gefunden, die sowohl ausreichende Vertraulichkeit als auch hinreichend gutes Analysepotenzial gewährleistet. Dies liegt insbesondere an den niedrigen Besetzungszahlen dieser Unternehmensgrößenklassen. Die beiden nachfolgenden Tabellen geben einen Überblick über die wichtigsten hochgerechneten Kenngrößen des Scientific-Use-Files, das noch aus 12 646 Unternehmen in der Stichprobe besteht. Während Tabelle 1 einen Gesamtüberblick liefert, zeigt Tabelle 2 die Verteilung der Unternehmen, deren Umsatz und Beschäftigte auf die Wirtschaftszweige.

Tabelle 1: Unternehmen, Beschäftigte, Umsatz, Rohertrag und Investitionen des Einzelhandels 1999

Datensatz	Unternehmen	Beschäftigte	Umsatz	Roh-ertrag <sup>1)</sup>	Roh-ertrags- quote <sup>2)</sup>	Investitionen
	Anzahl		Mill. EUR		%	Mill. EUR
Scientific-Use-File	108 312	889 293	101 181	33 764	33,4	1 405
Monatsmelder	110 921	2 114 018	284 758	93 812	32,9	4 616

1) Unternehmen mit fehlenden Werten bei der Berechnung des Rohertrages sind nicht berücksichtigt. - 2) Rohertragsquote = Rohertrag : Umsatz.

Demnach repräsentieren die im Scientific-Use-File der Einzelhandelsstatistik 1999 enthaltenen Unternehmen 97,6% der Grundgesamtheit aller Unternehmen mit einem Mindestjahresumsatz von 250 000 Euro und tragen zu gut einem Drittel zum Gesamtumsatz im Einzelhandel bei.

## 2 Anonymisierungsmaßnahmen

### 2.1 Traditionelle Verfahren

Die Daten für das Scientific-Use-File der Einzelhandelsstatistik 1999 wurden formal anonymisiert, das heißt direkte Identifikatoren wie Name und Anschrift wurden entfernt.

Tabelle 2: Unternehmen, Umsatz und Beschäftigte des Scientific-Use-Files der Einzelhandelsstatistik 1999 nach Wirtschaftszweigen

Wirtschaftszweig	WZ 93 <sup>1)</sup>	Unternehmen		Umsatz		Beschäftigte	
		Anzahl	%	Mill. EUR	%	Anzahl	%
Einzelhandel mit Waren verschiedener Art (in Verkaufsräumen) .....	52.1	13 980	12,9	13 785	13,6	139 872	15,7
Facheinzelhandel mit Nahrungsmitteln, Getränken und Tabakwaren (in Verkaufsräumen) .....	52.2	11 912	11,0	7 940	7,8	85 866	9,7
Apotheken; Facheinzelhandel mit medizinischen, orthopädischen und kosmetischen Artikeln (in Verkaufsräumen) .....	52.3	20 960	19,4	24 144	23,9	174 490	19,6
Einzelhandel mit Textilien .....	52.41	946	0,9	634	0,6	8 310	0,9
Einzelhandel mit Bekleidung .....	52.42	9 870	9,1	7 426	7,3	86 985	9,8
Einzelhandel mit Schuhen und Lederwaren .....	52.43	3 457	3,2	2 349	2,3	29 859	3,4
Einzelhandel mit Möbeln, Einrichtungs- gegenständen und Hausrat .....	52.44	7 523	6,9	8 614	8,5	70 321	7,9
Einzelhandel mit elektrischen Haushaltsgeräten, Geräten der Unterhaltungselektronik und Musikinstrumenten .....	52.45	6 168	5,7	5 334	5,3	44 610	5,0
Einzelhandel mit Metallwaren, Anstrichmitteln, Bau- und Heimwerkerbedarf .....	52.46	3 350	3,1	4 888	4,8	32 257	3,6
Einzelhandel mit Büchern, Zeitschriften, Zeitungen, Schreibwaren und Bürobedarf .....	52.47	4 102	3,8	2 876	2,8	30 497	3,4
Facheinzelhandel, anderweitig nicht genannt (in Verkaufsräumen) .....	52.48	17 223	15,9	12 498	12,4	132 379	14,9
Einzelhandel mit Antiquitäten und Gebrauchsgütern (in Verkaufsräumen) .....	52.5	254	0,2	236	0,2	1 855	0,2
Einzelhandel (nicht in Verkaufsräumen) .....	52.6	7 677	7,1	9 933	9,8	45 975	5,2
Reparatur von Gebrauchsgütern .....	52.7	890	0,8	524	0,5	6 017	0,7
Insgesamt ...	52	108 312	100	101 181	100	889 293	100

1) Klassifikation der Wirtschaftszweige, Ausgabe 1993.

Die Merkmale aus dem Bereich „Umsatz nach Tätigkeiten bzw. Produkten (in %)“ wurden herausgenommen, da die Information über die Nebentätigkeiten eines Unternehmens die Reidentifikationsgefahr stark ansteigen lässt. Dies konnte anhand der kommerziell zugänglichen MARKUS-Datenbank<sup>3)</sup> festgestellt werden. Die Information zur Haupttätigkeit eines Unternehmens bleibt durch die Kenntnis des Wirtschaftszweigs erhalten.

Das Merkmal „Anzahl der rechtlich unselbstständigen örtlichen Einheiten des Unternehmens“ (Filialen) kann von einem Datenangreifer als Überschneidungsmerkmal benutzt werden. Daher wurden für die reidentifikationsgefährdeten größeren Unternehmen bestimmte Kategorien für die Anzahl ihrer Filialen gebildet.

Bei der Regionalangabe wurde auf Wunsch des Wissenschaftlichen Begleitkreises analog zum Vorgehen bei der Erstellung der Scientific-Use-Files der Kostenstrukturerhebung 1999 und der Umsatzsteuerstatistik 2000 die Ost-West-Klassifizierung verwendet.<sup>4)</sup>

Die Wirtschaftszweigklassifikation wurde vom Fünfsteller auf den Dreisteller vergrößert. Ausnahme bildet der Dreisteller „52.4“ [„Sonstiger Facheinzelhandel (in Verkaufsräumen)“], der aufgrund seiner großen Besetzungszahl auf der Vierstellerebene angegeben werden kann. Dadurch bleiben dem Datennutzer inhaltlich wertvolle Informationen erhalten (siehe Tabelle 2).

## 2.2 Datenverändernde Verfahren

Abgesehen von den im nächsten Abschnitt beschriebenen punktuellen Maßnahmen bei bestimmten Unternehmen wurden generell keine datenverändernden Verfahren angewendet.

## 2.3 Zusätzliche Verfahren

Um eine faktische Anonymität zu gewährleisten, wurden schließlich noch drei „punktuell“ Maßnahmen – eine traditionelle Form der Anonymisierung und zwei datenverändernde Verfahren – angewandt. Davon waren lediglich sieben Unternehmen (dies entspricht rund 0,05% der Unternehmen aus der Stichprobe) aus den beiden relativ dünn besetzten Wirtschaftszweigen „52.5“ [„Einzelhandel mit Antiquitäten und Gebrauchsgütern (in Verkaufsräumen)“] und „52.41“ [„Einzelhandel mit Textilien“] betroffen.

Die Datensatzbeschreibung für die anonymisierten Daten enthält die Übersicht.

## 3 Schutz und Analysepotenzial

Zur Messung der Schutzwirkung wurden im Statistischen Bundesamt sowohl Simulationsexperimente von Massenfischzügen als auch von Einzelangriffen durchgeführt. Während ein Datenangreifer bei den Massenfischzügen mittels

3) Informationen zur Datenbank MARKUS finden sich auf der Internetseite [http://www.branchen-adressen.de/html/markus\\_marketingdatenbank.html](http://www.branchen-adressen.de/html/markus_marketingdatenbank.html). Siehe auch Lenz, R./Sturm, R./Vorgriemer, D.: „Maße für die faktische Anonymität von Mikrodaten“ in WiSta 6/2004, S. 621 ff., hier: S. 623.

4) Siehe Lenz, R.: „Ein Scientific-Use-File der Kostenstrukturerhebung im Verarbeitenden Gewerbe“ in WiSta 2/2005, S. 91 ff. bzw. Vorgriemer, D./Dittrich, S./Lenz, R./Rosemann, M.: „Ein Scientific-Use-File der Umsatzsteuerstatistik 2000“, in diesem Heft, S. 201 ff.

eines geeigneten Matchingverfahrens<sup>5)</sup> versucht, so viele Unternehmen wie möglich zu identifizieren, ist bei einem Einzelangriff<sup>6)</sup> ein bestimmtes Unternehmen von Interesse. In Tabelle 3 sind die aus verschiedenen Massenfischzugexperimenten<sup>7)</sup> berechneten Reidentifikationsrisiken für die neuen Länder und Berlin sowie das frühere Bundesgebiet (ohne Berlin) nach zwei Beschäftigtengrößenklassen dargestellt.

Tabelle 3: Reidentifikationsrisiken für Unternehmen nach Beschäftigtengrößenklassen  
Prozent

Unternehmenssitz	Insgesamt	Unternehmen mit ... bis ... Beschäftigten	
		1 – 19	20 – 49
Früheres Bundesgebiet (ohne Berlin) .....	12,5	12,3	13,8
Neue Länder und Berlin .....	15,8	15,2	21,5
Deutschland insgesamt ...	13,4	13,1	15,9


Die Tabelle 3 zeigt, dass für Deutschland nur rund 13% der Unternehmen reidentifiziert werden konnten. Bezieht man die Auswahlsätze<sup>8)</sup> der Unternehmen aus der Stichprobe mit ein, so konvergiert die Wahrscheinlichkeit, dass ein Unternehmen gefunden wird, gegen null.

Durch den Verzicht<sup>9)</sup> auf datenverändernde Anonymisierungsverfahren erhält der Datennutzer die Originalwerte der Erhebungsdaten. Einschränkungen der Auswertungsmöglichkeiten ergeben sich daher lediglich durch die im vorigen Abschnitt erläuterten traditionellen Anonymisierungsmaßnahmen. Beispielsweise sind Regionalanalysen „nur“ nach Ost und West möglich. Sofern Wissenschaftler/-innen an anderen Regionalanalysen interessiert sind, können sie über die Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder verschiedene Zugangsmöglichkeiten dafür beantragen.

#### 4 Schlussbemerkung

Neben dem Scientific-Use-File der Einzelhandelsstatistik 1999 stellen die faktisch anonymen Mikrodatenfiles der Kostenstrukturerhebung 1999<sup>10)</sup> und der Umsatzsteuerstatistik 2000<sup>11)</sup> weitere Datenangebote für die Wissenschaft dar, die bereits während des laufenden Projekts „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ entwickelt wurden. Gegenüber den anderen beiden Statistiken konnte bei der Generierung des Scientific-Use-Files im Falle der Einzelhandelsstatistik auf datenverändernde Verfahren nahezu verzichtet werden. Begünstigt wurde dies insbeson-

dere durch die große Anzahl an Unternehmen mit weniger als 50 Beschäftigten. Der Datennutzer bekommt hier die Möglichkeit, mit unverfälschten Daten an dem von ihm bevorzugten Ort arbeiten zu können. Mit der vorgestellten Datei erhält der Wissenschaftler Informationen über 97,6% der Grundgesamtheit<sup>12)</sup> im Einzelhandel. Damit kann ein guter Einblick in die Struktur des Einzelhandels gewonnen werden. Der Zugang zu den relativ wenigen großen Unternehmen ist über das Fernrechnen oder eine speziell formulierte Nutzeranfrage möglich.

Die Scientific-Use-Files sind für jeweils 65,- Euro 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 <http://www.forschungsdatenzentren.de> zur Verfügung. 

5) Eine theoretische Beschreibung des Verfahrens findet sich in Lenz, R.: "A graph theoretical approach to record linkage", Monographs of official statistics – Work session on statistical data confidentiality, 2003, S. 324 ff.

6) Siehe Vorgrimler, D.: „Re-Identifikationsmöglichkeiten am Beispiel eines konkreten Angriffsszenarios“ in Ronning, G./Gnoss, R. (Hrsg.): „Anonymisierung wirtschaftsstatistischer Einzeldaten“, Band 42 der Schriftenreihe „Forum der Bundesstatistik“, Wiesbaden 2003, S. 40 ff.

7) Zum genaueren Vorgehen siehe Lenz, R./Sturm, R./Vorgrimler, D., a. a. O. (Fußnote 3).

8) Beispielsweise beträgt der kleinste Auswahlsatz rund 1,4%.

9) Die wenigen Ausnahmen sind im Abschnitt 2.3 beschrieben.

10) Siehe Fußnote 4.

11) Siehe Fußnote 4.

12) Bezogen auf die Unternehmen mit einem Mindestjahresumsatz von 250 000 Euro.

*Dr. Daniel Vorgrimler, Dipl.-Volkswirt Stefan Dittrich, Dr. Rainer Lenz (alle Statistisches Bundesamt),  
Dipl.-Volkswirt Martin Rosemann (Institut für angewandte Wirtschaftsforschung – IAW, Tübingen)*

# Ein Scientific-Use-File der Umsatzsteuerstatistik 2000

*Nachdem im Sommer 2004 mit dem Scientific-Use-File der Lohn- und Einkommensteuerstatistik 1998<sup>1)</sup> erstmalig Einzeldaten der Steuerstatistiken der Wissenschaft in standardisierter Form zur Verfügung gestellt werden konnten, veröffentlicht nun die deutsche amtliche Statistik mit dem Scientific-Use-File zur Umsatzsteuerstatistik 2000 ein weiteres Mikrodatenfile aus den Steuerstatistiken. Wie bei der Lohn- und Einkommensteuerstatistik können Wissenschaftler mit diesen Mikrodaten unter den Prämissen des § 16 Abs. 6 Bundesstatistikgesetz (BStatG)<sup>2)</sup> am eigenen Arbeitsplatz wissenschaftliche Analysen durchführen.*

*Die Grundlagen zur Sicherstellung der faktischen Anonymität nach § 16 Abs. 6 BStatG bei gleichzeitigem Erhalt des Analysepotenzials wurden im Projekt „Faktische Anonymisierung wirtschaftsstatistischer Einzeldaten“ erarbeitet. Weitere Scientific-Use-Files, die dank der Arbeiten dieses Projektes veröffentlicht werden können, basieren auf Daten der Kostenstrukturerhebung im Verarbeitenden Gewerbe und der Einzelhandelsstatistik.<sup>3)</sup>*

*Im folgenden Beitrag wird zunächst die Umsatzsteuerstatistik als solche vorgestellt. Das zweite Kapitel beschreibt das Anonymisierungskonzept. Mit den Möglichkeiten, welche die Umsatzsteuerstatistik und das Scientific-Use-File bieten, und mit den Einschränkungen bei den Analysen aufgrund*

*der Anonymisierung beschäftigt sich das dritte Kapitel. Ein Fazit rundet den Beitrag ab.*

## 1 Die Umsatzsteuerstatistik

### 1.1 Datengrundlage

Im Rahmen der Umsatzbesteuerung werden von den Unternehmen Umsatzsteuer-Voranmeldungen bei den Finanzbehörden abgegeben.<sup>4)</sup> Durch die Auswertung der monatlichen und vierteljährlichen Umsatzsteuer-Voranmeldungen ergänzt um Daten des Grundinformationsdienstes<sup>5)</sup> gewinnt die amtliche Statistik Informationen über die Entstehung der Umsatzsteuer, über steuerpflichtige Unternehmen und deren Umsätze sowie über die innergemeinschaftlichen Erwerbe [Einfuhren aus anderen Mitgliedstaaten der Europäischen Union (EU)]. Durch die relativ kurzen Abgabefristen der Voranmeldungen liegen bereits etwa 14 Monate nach Ablauf eines Kalenderjahres Angaben über die Entwicklung der Umsätze in fast allen Wirtschaftsbereichen vor. Abweichungen zwischen den Angaben der Voranmeldungen und den tatsächlichen Umsätzen (beispielsweise durch Korrekturen bei Warenrückgaben) sind möglich, jedoch insgesamt nicht von größerer Bedeutung. Ein weiterer Vorteil der

1) Siehe Merz, J./Vorgrimler, D./Zwick, M.: „Faktisch anonymisiertes Mikrodatenfile der Lohn- und Einkommensteuerstatistik 1998“ in WiSta 10/2004, S. 1079 ff.

2) Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) vom 22. Januar 1987 (BGBl. I S. 462, 565), zuletzt geändert durch Artikel 16 des Gesetzes vom 21. August 2002 (BGBl. I S. 3322).

3) Für die Kostenstrukturerhebung siehe Lenz, R./Vorgrimler, D./Rosemann, M.: „Ein Scientific-Use-File der Kostenstrukturerhebung im Verarbeitenden Gewerbe“ in WiSta 2/2005, S. 91 ff. Für die Einzelhandelsstatistik siehe Scheffler, M.: „Ein Scientific-Use-File der Einzelhandelsstatistik 1999“ in diesem Heft, S. 197 ff.

4) Neben den Umsatzsteuer-Voranmeldungen werden jährliche Umsatzsteuererklärungen abgegeben, die der endgültigen Steuerfestsetzung (Veranlagung) dienen. Die Voranmeldepflicht entfällt bei Unternehmen, die im vorangegangenen Jahr höchstens 511 Euro Umsatzsteuer zu entrichten hatten (sog. Jahreszahler).

5) Im Grundinformationsdienst werden von der Finanzverwaltung Angaben zu den persönlichen Verhältnissen der Steuerpflichtigen gespeichert, z. B. Wohnort oder Gewerkekennzahl.

Umsatzsteuerstatistik ist die durch die Ausgestaltung als Sekundärstatistik weniger starke Belastung der Wirtschaftseinheiten im Vergleich zu einer direkten Befragung. Die Durchführung der Umsatzsteuerstatistik ist im Gesetz über Steuerstatistiken (StStatG) vom Oktober 1995<sup>6)</sup> geregelt.

Umsatzsteuerpflichtig und damit in der Umsatzsteuerstatistik abgebildet sind grundsätzlich alle Unternehmer, die Lieferungen und Leistungen im Inland gegen Entgelt im Rahmen ihres Unternehmens ausführen. Nicht erfasst sind in der Umsatzsteuerstatistik Unternehmen, die aufgrund ihrer Umsatzhöhe oder durch die Erbringung vorwiegend umsatzsteuerfreier Leistungen von der unterjährigen Abgabe von Umsatzsteuer-Voranmeldungen befreit sind (z. B. niedergelassene Ärzte, Behörden, Theater und Museen). Die Umsatzsteuerstatistik liefert somit für fast alle Wirtschaftsbereiche Daten; bei Vergleichen zwischen verschiedenen Wirtschaftsbereichen sind allerdings Untererfassungen aufgrund steuerlicher Besonderheiten zu berücksichtigen. Weiterhin ist zu beachten, dass bei finanziell, wirtschaftlich und organisatorisch zusammengehörenden Unternehmen (z. B. Filialen, Zweigbetrieben oder Tochterunternehmen) die Unternehmen als Einheit mit dem gesamten Jahresumsatz von dem für den Sitz der Geschäftsleitung zuständigen Finanzamt zentral erfasst werden. Besteuert und somit in der Statistik abgebildet werden lediglich die Außenumsätze des Unternehmenskreises. Das Fehlen der Innenumsätze des Unternehmenskreises unterzeichnet somit die Höhe der Unternehmensumsätze in der Umsatzsteuerstatistik. Auch für Untersuchungen nach Wirtschaftszweigen oder Regionen ist die Zusammenfassung der Umsätze eines Unternehmenskreises als Nachteil zu bewerten, da gegebenenfalls Unternehmensteile unterschiedlicher Tätigkeitsfelder bzw. unterschiedlicher Regionen zusammengefasst werden.

## 1.2 Aussagekraft der Umsatzsteuerstatistik

In der Umsatzsteuerstatistik werden u. a. folgende qualitative Merkmale abgebildet: Gewerbekennzahl, amtlicher Gemeindeschlüssel, letztgültige Zahlungsweise, Dauer der Steuerpflicht, Organschaft, Rechtsform. Die quantitativen Merkmale umfassen u. a. steuerbaren Umsatz (ohne die der Einfuhrumsatzsteuer unterliegenden Umsätze), Umsatzsteuer vor Abzug der Vorsteuerbeträge, abziehbare Vorsteuerbeträge, Umsatzsteuer-Vorauszahlung. Für wirtschaftsstatistische Untersuchungen ist dabei der Umsatz der Unternehmen von besonderem Interesse. Zum steuerbaren Umsatz der Unternehmen zählen im Umsatzsteuerrecht neben den als „Lieferungen und Leistungen“ bezeichneten Umsätzen der Unternehmen auch die von ihnen aus EU-Ländern importierten Güter („innergemeinschaftliche Erwerbe“). Als Umsatz im wirtschaftsstatistischen Sinn werden aber lediglich die Lieferungen und Leistungen betrachtet, da die innergemeinschaftlichen Erwerbe den Vorleistungen zuzurechnen sind. Im Vergleich zu Primärerhebun-

gen ist neben der erhebungstechnischen Abgrenzung des Umsatzes auch die inhaltliche Abgrenzung des steuerlichen Umsatzbegriffes zu beachten. Während in Primärerhebungen allein der Umsatz aus der laufenden Produktionstätigkeit erfragt wird, umfasst der steuerliche Umsatz zusätzlich außerordentliche Erträge, zum Beispiel aus einem nicht dem Betriebszweck dienenden Verkauf von Gebäuden.<sup>7)</sup>

Das am häufigsten genutzte qualitative Merkmal der Umsatzsteuerstatistik ist die Wirtschaftszweiguordnung. Maßgebend für die Zuordnung zu einem Wirtschaftszweig ist der Schwerpunkt der wirtschaftlichen Tätigkeit eines Unternehmens. Seit dem Berichtsjahr 2002 liegt der Branchengliederung der Umsatzsteuerstatistik die aktualisierte Ausgabe 2003 der Klassifikation der Wirtschaftszweige (WZ 2003) in der für die Steuerstatistiken gültigen Tiefengliederung zu Grunde (GKZ 2003)<sup>8)</sup>. Für das Berichtsjahr 2000, für das das Scientific-Use-File vorliegt, war die Ausgabe 1993 (WZ 93) relevant.

Ergebnisse der Umsatzsteuerstatistik zu allen 830 nachgewiesenen Wirtschaftszweigen können kostenlos über den Statistik-Shop des Statistischen Bundesamtes bezogen werden ([www.ec-destatis.de](http://www.ec-destatis.de)).<sup>9)</sup>

## 2 Anonymisierung

Um die zur Weitergabe an die Wissenschaft notwendige faktische Anonymität der Merkmalsträger nach § 16 Abs. 6 BStatG zu gewährleisten, mussten Informationen aus dem Originalmaterial der Umsatzsteuerstatistik vergrößert, gelöscht und in wenigen Fällen verändert werden. Der Einsatz von datenverändernden Verfahren (in diesem Fall der Mikroaggregation) wurde hierbei auf ein Minimum beschränkt. Im Folgenden wird die Anonymisierungskonzeption zusammen mit ihrer Schutzwirkung vorgestellt.

### 2.1 Anonymisierungskonzeption

Die Umsatzsteuerstatistik umfasst wie erwähnt qualitative und quantitative Merkmale. Die qualitativen Merkmale sind für den Datenschutz besonders dann problematisch, wenn sie aufgrund einer tiefen Gliederung die Merkmalsträger in einer solchen Weise beschreiben, dass nur noch wenige Unternehmen dieser Beschreibung entsprechen (z. B. der einzige Bäcker einer Gemeinde).

Aus diesem Grunde sieht das Anonymisierungskonzept bei den qualitativen Merkmalen folgende Maßnahmen vor:

- Die *Regionalangabe* wird auf Ost/West vergrößert, wobei der Osten die neuen Bundesländer einschließlich Berlin umfasst.

6) Artikel 35 des Jahressteuergesetzes 1996 vom 11. Oktober 1995 (BGBl. I S. 1250).

7) Siehe Treck, H.-J.: „Die Umsatzsteuerstatistik als Quelle wirtschaftsstatistischer Analysen“ in Statistische Analysen und Studien Nordrhein-Westfalen, Band 15, S. 3 ff.

8) Die Klassifikation der Gewerbekennziffern (GKZ) entspricht in ihrem Aufbau der Klassifikation der Wirtschaftszweige. Aus Geheimhaltungs- und Praktikabilitätsgründen werden aber nicht alle Positionen der WZ in der tiefsten Gliederung nachgewiesen.

9) Siehe auch Dittrich, S.: „Umsätze und ihre Besteuerung 2002“ in WiSta 10/2004, S. 1195 ff.



- Die *Wirtschaftszweiguordnung* wird in unterschiedlicher Tiefengliederung abhängig von den Besetzungszahlen in die Daten aufgenommen (zur Einteilung siehe die Anhangtabelle auf S. 210). Dabei werden zum Teil auch neue zusammengefasste Positionen gebildet.
- Aus dem Merkmal *Beginn der Steuerpflicht* wird das Merkmal *Neugründung* mit den Ausprägungen 1 = ja und 0 = nein gebildet. Bei Unternehmen mit mehr als 100 Mill. Euro Umsatz wird das Merkmal generell auf 0 gesetzt.<sup>10)</sup>
- Das Merkmal *Rechtsform* wird vergrößert (mit den Ausprägungen Personengesellschaften, Kapitalgesellschaften, Genossenschaften und sonstige Rechtsformen).

Bei den quantitativen Merkmalen (z.B. Umsatz) entstehen besonders bei extremen Ausprägungen datenschutzrechtliche Probleme. So ist zum Beispiel das Unternehmen mit dem höchsten Umsatz in besonderer Weise gefährdet. Verschärft wird dieses Problem bei einer Kombination von quantitativen und qualitativen Merkmalen. Wird zum Beispiel der Umsatz innerhalb eines Wirtschaftszweiges bzw. einer Branche betrachtet, ist der jeweilige Branchenmarktführer besonders gefährdet.<sup>11)</sup> Daher besteht das Anonymisierungskonzept bei den quantitativen Merkmalen aus einer *zweistufigen Mikroaggregation*. Die erste Stufe beinhaltet eine für jedes Merkmal getrennt durchgeführte Mikroaggregation für alle Unternehmen.<sup>12)</sup> Dies gilt als die schwächste Form der Mikroaggregation.<sup>13)</sup> Da durch diese Maßnahme die regionalen Branchenmarktführer<sup>14)</sup> noch nicht genügend geschützt werden (siehe Abschnitt 2.2), wird als zweite Stufe eine punktuelle Mikroaggregation durchgeführt. Dabei werden nur speziell die jeweiligen drei regionalen Marktführer einer Branche mikroaggregiert, wobei das Merkmal *Lieferungen und Leistungen* (Umsatz) das bestimmende Merkmal ist. Da in dem Scientific-Use-File zwischen 68 Wirtschaftszweigen und zwei Regionen unterschieden wird, sind insgesamt 408 Merkmalsträger von dieser zweiten Stufe betroffen (jeweils drei Unternehmen in 136 regionalisierten Wirtschaftszweigen).

Je herausragender eine Ausprägung ist, desto stärker weicht der mikroaggregierte Wert vom Originalwert ab. Bei kleineren Unternehmen unterscheiden sich die Umsätze kaum. Bei einem Großunternehmen unterscheidet sich hingegen der Umsatz eventuell deutlich vom Umsatz des nächstgrößeren Unternehmens. Daher werden bei der Mikroaggregation die Merkmalsausprägungen der kleinen und mittleren Unternehmen nur minimal verändert, die der großen hingegen stärker. Je nach Merkmal sind zwischen 90 und 99% der Ausprägungen trotz der Mikroaggregation unverändert in das Scientific-Use-File eingegangen. Nur sehr wenige

Merkmalsausprägungen weisen relative Abweichungen von mehr als 5% zwischen mikroaggregiertem und originalelem Wert auf. Betrachtet man jedoch lediglich die Unternehmen mit mehr als 500 Mill. Euro Umsatz, dann ist gut die Hälfte der Ausprägungen durch die Anonymisierung verändert worden, allerdings etwa 90% innerhalb eines engen Bereichs von 5% relativer Abweichung. Da die Großunternehmen hinsichtlich einer Reidentifikation gefährdeter sind, erfüllt diese Art der Anonymisierung die Forderung, Merkmalsträger abhängig vom Grad ihres Reidentifikationsrisikos zu anonymisieren.

## 2.2 Schutzwirkung der Anonymisierung

In den Beiträgen von Lenz/Sturm/Vorgrimler sowie Lenz/Vorgrimler in der Ausgabe 6/2004<sup>15)</sup> dieser Zeitschrift wurde die Schutzwirkung der Anonymisierung für die Umsatzsteuerstatistik bereits ausführlich dargestellt, sodass an dieser Stelle einige zusätzliche Anmerkungen genügen.

Die zitierten Arbeiten zur Schutzwirkung zeigten in zwei Bereichen weiteren Forschungs- bzw. Anonymisierungsbedarf auf. Da in diesen Arbeiten Dateien zu Grunde gelegt wurden, bei denen die Marktführer nicht genügend geschützt wurden, bestand zum einen ein zusätzlicher Anonymisierungsbedarf bei den jeweiligen Marktführern. Zum anderen bestand ein methodischer Forschungsbedarf, da sich die Schutzanalysen lediglich auf einen kleinen Ausschnitt der in der Umsatzsteuerstatistik enthaltenen Unternehmen beschränkten (Unternehmen aus dem Verarbeitenden Gewerbe mit mindestens 20 Beschäftigten).

Wie in Abschnitt 2.1 beschrieben wurde zum zusätzlichen Schutz der Marktführer die punktuelle Mikroaggregation eingesetzt. Diese führt dazu, dass sich die drei regionalen Branchenmarktführer mit Ausnahme der Rechtsform nicht mehr unterscheiden. Eine eindeutige Zuordnung ist nur noch dann möglich, wenn sich einer der drei Merkmalsträger durch seine Rechtsform eindeutig hervorhebt. Bei den 136 regionalen Wirtschaftszweigen (jeweils 68 Wirtschaftszweige im Osten und im Westen) unterscheidet sich in 95 Fällen mindestens eines der drei marktführenden Unternehmen von den anderen durch die Rechtsform und könnte somit theoretisch zugeordnet werden. Eine solche eindeutige Zuordnung würde aber nicht gegen die faktische Anonymität verstoßen, da ein Datenangreifer keine zusätzlichen Informationen gewinnen kann. So würde er zum Beispiel auch nach einer korrekten Zuordnung weiterhin lediglich den durchschnittlichen Umsatz der drei regionalen Branchenmarktführer kennen und nicht den exakten Umsatzwert seines zugeordneten Unternehmens. Dasselbe gilt für die restlichen im Datensatz vorhandenen quantitativen Merk-

10) Von über 150 000 als Neugründungen gekennzeichneten Unternehmen haben 118 Unternehmen einen Umsatz von über 100 Mill. Euro. Bei diesen Unternehmen wird das Merkmal „Neugründung“ auf 0 gesetzt und damit die Information unterdrückt. Unter Plausibilitäts Gesichtspunkten dürfte diese Informationsreduktion nicht besonders relevant sein, da es sich in der Mehrheit der Fälle um keine echten Neugründungen handeln wird. Zu „unechten“ Neugründungen kommt es beispielsweise bei Rechtsformänderungen oder Sitzverlagerungen.

11) Siehe Lenz, R./Sturm, R./Vorgrimler, D.: „Maße für die faktische Anonymität von Mikrodaten“ in WiSta 6/2004, S. 628 f. und S. 636 f.

12) Zum Verfahren der Mikroaggregation siehe Höhne, J.: „Methoden zur Anonymisierung wirtschaftsstatistischer Einzeldaten“ in Ronning, G./Gnoss, R. (Hrsg.): „Anonymisierung wirtschaftsstatistischer Einzeldaten“, Band 42 der Schriftenreihe „Forum der Bundesstatistik“, Wiesbaden 2003, S. 77 ff.

13) Zur Wirkung der Mikroaggregation bei der Umsatzsteuerstatistik siehe Lenz, R./Vorgrimler, D.: „Geheimhaltungsmethoden auf dem Prüfstand – eine Analyse anhand der Umsatzsteuerstatistik“ in WiSta 6/2004, S. 639 ff.

14) Mit den regionalen Branchenmarktführern sind die jeweils drei Unternehmen gemeint, die in den jeweiligen Branchen (abgegrenzt nach der Wirtschaftsklassifikation) innerhalb ihrer Region (Ost/West) die höchsten Umsätze erzielen.

15) Siehe Fußnoten 11 und 13.

male. Die durchschnittlichen Werte liegen darüber hinaus in gut der Hälfte aller Fälle relativ um wenigstens 50% von den originalen Werten entfernt. Die einzige bei einer korrekten Zuordnung zu gewinnende zusätzliche Information wäre daher die Rechtsform der Unternehmen. Diese Information müsste aber ein Datenangreifer für den Reidentifikationsversuch einsetzen. Er könnte sie also nicht aus einem „Datenangriff“ gewinnen. Die Gruppe der regionalen Branchenmarktführer kann daher als faktisch anonym angesehen werden.

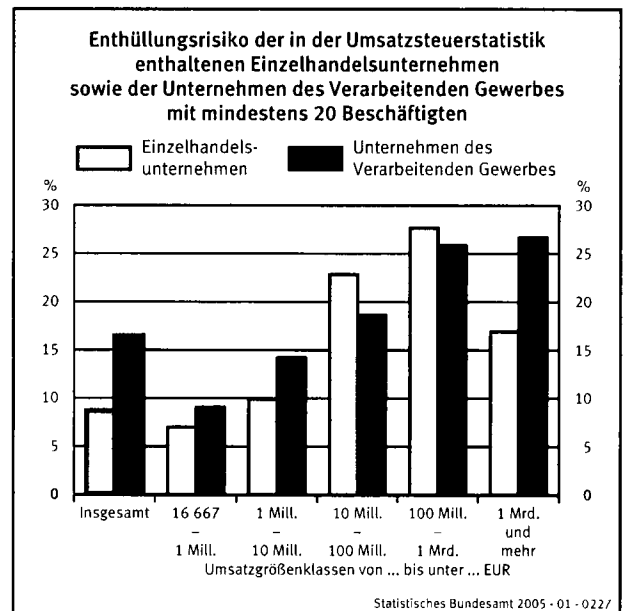
Um die eingeschränkte Sichtweise bei der Schutzwirkungsanalyse um einen weiteren Wirtschaftsbereich und um Kleinunternehmen zu erweitern, wurde ein weiterer Massenfischzug simuliert, bei dem die Daten der Einzelhandelsstatistik als Zusatzwissen verwendet wurden.<sup>16)</sup> Strukturell unterscheidet sich diese Untersuchung von den vorhergehenden in zweierlei Hinsicht:

- in der geringeren Anzahl unterschiedlicher Wirtschaftszweige (7 gegenüber 22),
- in der fehlenden „Abschneidegrenze“ für kleine Unternehmen.

Beide Punkte sprechen für wesentlich schlechtere Voraussetzungen für einen Massenfischzug mit den Daten der Einzelhandelsstatistik als mit den Daten der Kostenstrukturerhebung.

Insgesamt wurden 12500 Unternehmen der Einzelhandelsstatistik unter mehr als 300000 Unternehmen aus der Umsatzsteuerstatistik gesucht (bei den Unternehmen des Verarbeitenden Gewerbes mit mindestens 20 Beschäftigten waren es rund 9000 gesuchte Unternehmen unter etwas mehr als 37000 Unternehmen). Die benötigte Rechenlaufzeit war mit 19 Stunden CPU-Zeit außerordentlich hoch.

Das Schaubild zeigt die berechneten Enthüllungsrisiken<sup>17)</sup>. Zum Vergleich sind neben den Risiken, die sich mit der Einzelhandelsstatistik ergeben, auch die Enthüllungsrisiken dargestellt, die mit der Kostenstrukturerhebung berechnet wurden. Die jetzt berechneten Ergebnisse für die Unternehmen des Verarbeitenden Gewerbes stimmen nicht ganz mit den früher veröffentlichten Ergebnissen überein, da in diesen die bereits erwähnte punktuelle Mikroaggregation nicht berücksichtigt wurde und sich die Gliederungstiefe der Wirtschaftsklassifizierung leicht verändert hat. Für die Unternehmen des Einzelhandels ergibt sich ein insgesamt deutlich geringeres Enthüllungsrisiko als für die Unternehmen des Verarbeitenden Gewerbes. Dies liegt an der deutlich unterschiedlichen Größenstruktur der Unternehmen, die mit der fehlenden Abschneidegrenze in der Einzelhandelsstatistik zusammenhängt. Tabelle 1 zeigt die Verteilung der gesuchten Unternehmen auf die Umsatzgrößenklassen bei den beiden Massenfischzügen. Bei den Einzelhandels-



unternehmen dominieren die Unternehmen der kleinsten Umsatzgrößenklasse, daher wirken sie sich auch dominierend auf das Maß für das Enthüllungsrisiko aus. Da dieses bei den kleinsten Unternehmen sehr gering ist, ist auch das gesamte Enthüllungsrisiko für die Unternehmen des Einzelhandels sehr gering.

Tabelle 1: Verteilung der in der Umsatzsteuerstatistik 2000 gesuchten Einzelhandelsunternehmen und Unternehmen des Verarbeitenden Gewerbes auf die Umsatzgrößenklassen

Umsatzgrößenklassen von ... bis unter ... EUR	Gesuchte Einzelhandelsunternehmen		Gesuchte Unternehmen des Verarbeitenden Gewerbes	
	Anzahl	%	Anzahl	%
16 667 – 1 Mill. ...	8166	65,3	440	4,7
1 Mill. – 10 Mill. ...	3693	29,5	4757	51,2
10 Mill. – 100 Mill. ...	498	4,0	3426	36,9
100 Mill. – 1 Mrd. ...	130	1,0	616	6,6
1 Mrd. und mehr ...	13	0,1	44	0,5
<b>Insgesamt ...</b>	<b>12 500</b>	<b>100</b>	<b>9 283</b>	<b>100</b>

Wird von einer Gesamtbetrachtung zu einer Betrachtung des Enthüllungsrisikos nach Größenklassen des Umsatzes übergegangen, so zeigt sich, dass sich die Ergebnisse der beiden Szenarien nicht wesentlich unterscheiden. Auffallend ist allerdings das deutlich geringere Risiko bei den größten Unternehmen des Einzelhandels. Dies äußert sich sowohl durch eine geringere Trefferquote (23 gegenüber 43% bei den Unternehmen des Verarbeitenden Gewerbes) als auch durch einen geringeren Anteil an nützlichen Informationen (55 gegenüber 61%).

Die Ergebnisse zeigen den großen Schutzeffekt, den die getroffenen Anonymisierungsmaßnahmen erzeugen. Die

<sup>16)</sup> Zum Konzept des Massenfischzuges siehe Lenz, R./Vorgriemer, D., Fußnote 13.

<sup>17)</sup> Das Enthüllungsrisiko ist definiert als Wahrscheinlichkeit, eine nützliche Information zu enthüllen. Es handelt sich somit um die gemeinsame Wahrscheinlichkeit für die Ereignisse „richtige Zuordnung“ und „nützliche Information“. Eine Information ist hierbei die Ausprägung eines Merkmals bei einem Merkmalsträger. Nützlich ist diese Information für einen Datenangreifer, wenn sie nur bis zu einem bestimmten Wert von der originalen Ausprägung abweicht (diese Nützlichkeitschwelle wurde bei der Umsatzsteuerstatistik mit 10% festgelegt). Bei datenverändernden Verfahren – wie der Mikroaggregation – können die anonymisierten von den originalen Werten abweichen. Eine ausführliche Betrachtung des Enthüllungsrisikos findet sich in Höhne, J./Sturm, R./Vorgriemer, D.: „Konzept zur Beurteilung der Schutzwirkung von faktischer Anonymisierung“ in WiSta 4/2003, S. 287 ff.

Wahrscheinlichkeit für einen Datenangreifer, nützliche Informationen zu erhalten, ist selbst mit sehr gutem Zusatzwissen (was in diesen Fällen angenommen wurde) gering. Viel eher wird er unbrauchbare oder überhaupt keine Informationen erhalten. Dies alles ist vor dem Hintergrund zu sehen, dass der Datenangreifer eine nützliche von einer unbrauchbaren Information nicht unterscheiden kann. Er „stochert“ sozusagen „im Nebel“. Die Daten können daher als faktisch anonym im Sinne des § 16 Abs. 6 BStatG angesehen werden.

### 3 Analysemöglichkeiten und Vergleich mit den Originaldaten

#### 3.1 Umsatzanalysen

##### Eckzahlen

Das kassenmäßige Aufkommen aus der Umsatzsteuer (ohne Einfuhrumsatzsteuer) betrug im Jahr 2000 107 Mrd. Euro, das aus der Einfuhrumsatzsteuer 34 Mrd. Euro. Diese beiden Steuerarten trugen damit 28% zum gesamten Steuerertrag bei. Die von den Unternehmen an die Finanzämter abzuführenden Umsatzsteuer-Vorauszahlungen, die sich aus methodischen Gründen vom kassenmäßigen Steuerertrag unterscheiden, beliefen sich im Jahr 2000 ebenfalls auf 107 Mrd. Euro.

Im Jahr 2000 haben 2,9 Mill. Unternehmen in Deutschland Umsatzsteuer-Voranmeldungen abgegeben und dabei Umsätze in Höhe von 4 153 Mrd. Euro gemeldet. Die mit dem allgemeinen Steuersatz von 16% besteuerten Umsätze betragen 2 942 Mrd. Euro, mit dem ermäßigten Steuersatz

von 7% wurden 414 Mrd. Euro besteuert. Die Ergebnisse des faktisch anonymisierten Datenfiles weichen trotz des oben beschriebenen Anonymisierungskonzepts bei den Eckzahlen nicht von den Ergebnissen der Umsatzsteuerstatistik ab.

Die häufigsten Auswertungen der Umsatzsteuerstatistik betreffen Auszahlungen der Steuerpflichtigen und der Lieferungen und Leistungen nach Wirtschaftszweigen und Umsatzgrößenklassen. Tabelle 2 zeigt eine Gegenüberstellung für die Anzahl der Steuerpflichtigen und ihre Lieferungen und Leistungen nach Wirtschaftszweigen. Während die Wirtschaftszweiguordnung durch die Anonymisierung nicht verändert wurde und damit die Fallzahlen unverändert geblieben sind, ergeben sich für die Umsätze leichte Veränderungen, die jedoch mit bis zu maximal 2,4% im Abschnitt G „Handel; Instandhaltung und Reparatur von Kraftfahrzeugen und Gebrauchsgütern“ gering sind. Auch die weiteren quantitativen Merkmale des faktisch anonymisierten Datenfiles weichen nur in wenigen Fällen in größerem Ausmaß von den Originalwerten ab.<sup>18)</sup>

Bei einer Betrachtung nach Umsatzgrößenklassen wird der Einfluss der Mikroaggregation auf die quantitativen Merkmale der Großunternehmen deutlich, während die Ergebnisse für die Vielzahl der kleinen und mittleren Unternehmen unverändert bleiben. Tabelle 3 zeigt die Anzahl der Steuerpflichtigen und deren Lieferungen und Leistungen nach Umsatzgrößenklassen. Die Ergebnisse für die Unternehmen mit einem Jahresumsatz unter 100 Mill. Euro weisen keine nennenswerten Unterschiede zwischen der Umsatzsteuerstatistik und dem faktisch anonymisierten Datenfile auf. Lediglich bei den Unternehmen mit einem Umsatz von mehr als 100 Mill. Euro werden durch das Anonymisierungskonzept leicht unterschiedliche Ergebnisse bewirkt.

Tabelle 2: Steuerpflichtige, Lieferungen und Leistungen nach Wirtschaftszweigen  
Gegenüberstellung der Ergebnisse der Umsatzsteuerstatistik 2000 und des faktisch anonymisierten Datenfiles

WZ 93 <sup>1)</sup>	Wirtschaftsgliederung	Steuerpflichtige <sup>2)</sup>			Abweichung	Lieferungen und Leistungen <sup>3)</sup>		Abweichung
		Originalwerte	anonymisierte Werte	%		Originalwerte	anonymisierte Werte	
		Anzahl				Mill. EUR		
A – O	Wirtschaftszweige insgesamt .....	2 909 150	2 909 150	0,0	4 152 927	4 152 927	±0,0	
A, B	Land- und Forstwirtschaft, Fischerei und Fischzucht ...	65 764	65 764	0,0	23 316	23 320	±0,0	
C	Bergbau und Gewinnung von Steinen und Erden .....	3 067	3 067	0,0	26 601	26 198	-1,5	
D	Verarbeitendes Gewerbe .....	2 918 85	2 918 85	0,0	1 514 702	1 481 318	-2,2	
E	Energie- und Wasserversorgung .....	10 035	10 035	0,0	144 216	142 630	-1,1	
F	Baugewerbe .....	323 116	323 116	0,0	218 928	218 962	±0,0	
G	Handel; Instandhaltung und Reparatur von Kraftfahrzeugen und Gebrauchsgütern .....	731 491	731 491	0,0	1 328 683	1 361 060	+2,4	
H	Gastgewerbe .....	251 865	251 865	0,0	53 288	53 292	±0,0	
I	Verkehr und Nachrichtenübermittlung .....	127 391	127 391	0,0	217 334	220 937	+1,7	
J	Kredit- und Versicherungsgewerbe .....	16 052	16 052	0,0	30 917	30 958	+0,1	
K	Grundstücks- und Wohnungswesen, Vermietung beweglicher Sachen, Erbringung von Dienstleistungen überwiegend für Unternehmen .....	760 671	760 671	0,0	458 162	457 425	-0,2	
L, M, N	Öffentliche Verwaltung, Verteidigung, Sozialversicherung, Erziehung und Unterricht, Gesundheits-, Veterinär- und Sozialwesen .....	62 901	62 901	0,0	43 247	43 346	+0,2	
O	Erbringung von sonstigen öffentlichen und persönlichen Dienstleistungen .....	264 912	264 912	0,0	93 532	93 481	-0,1	

1) Klassifikation der Wirtschaftszweige, Ausgabe 1993. – 2) Mit mehr als 16 617 Euro Jahresumsatz. – 3) Umsätze der Unternehmen ohne Umsatzsteuer.

18) Um dem Nutzer des Datenfiles die Bereiche mit signifikanten Datenveränderungen kenntlich zu machen, wird die CD-ROM, die das Scientific-Use-File der Umsatzsteuerstatistik 2000 enthält, entsprechende Vergleichsauswertungen einschließen.

Tabelle 3: Steuerpflichtige, Lieferungen und Leistungen nach Umsatzgrößenklassen  
Gegenüberstellung der Ergebnisse der Umsatzsteuerstatistik 2000 und des faktisch anonymisierten Datenfiles

Größenklassen der Lieferungen und Leistungen von ... bis unter ... EUR	Steuerpflichtige <sup>1)</sup>			Lieferungen und Leistungen <sup>2)</sup>		
	Originalwerte	anonymisierte Werte	Abweichung	Originalwerte	anonymisierte Werte	Abweichung
	Anzahl		%	Mill. EUR		%
16 617 – 50 000 .....	773 820	773 821	±0,0	24 278	24 278	±0,0
50 000 – 100 000 .....	568 170	568 167	±0,0	40 959	40 959	±0,0
100 000 – 250 000 .....	662 982	662 986	±0,0	105 968	105 969	±0,0
250 000 – 500 000 .....	357 106	357 101	±0,0	126 237	126 235	±0,0
500 000 – 1 Mill. ....	238 229	238 230	±0,0	167 420	167 419	±0,0
1 Mill. – 2 Mill. ....	143 908	143 910	±0,0	200 770	200 772	±0,0
2 Mill. – 5 Mill. ....	93 323	93 321	±0,0	286 776	286 773	±0,0
5 Mill. – 10 Mill. ....	34 524	34 517	±0,0	240 764	240 700	±0,0
10 Mill. – 25 Mill. ....	21 591	21 598	±0,0	331 512	331 719	-0,1
25 Mill. – 50 Mill. ....	7 786	7 769	+0,2	270 984	270 345	+0,2
50 Mill. – 100 Mill. ....	3 874	3 879	-0,1	268 392	268 565	-0,1
100 Mill. – 250 Mill. ....	2 352	2 346	+0,3	359 880	358 363	+0,4
250 Mill. und mehr .....	1 485	1 505	-1,3	1 728 987	1 730 831	-0,1
Insgesamt ...	2 909 150	2 909 150	±0,0	4 152 927	4 152 927	±0,0

1) Mit mehr als 16 617 Euro Jahresumsatz. – 2) Umsätze der Unternehmen ohne Umsatzsteuer.

### 3.2 Konzentrationsanalysen

Eine nahe liegende Untersuchungsmöglichkeit mit den Ergebnissen der Umsatzsteuerstatistik ist die Analyse der Verteilung der Umsätze auf die Unternehmen und daraus abgeleitet die Untersuchung der Unternehmenskonzentration. Konzentration lässt sich auf zweierlei Arten messen<sup>19)</sup>:

- als statische Konzentration, das heißt Konzentration als bestehender Zustand, und
- als dynamische Konzentration, das heißt Konzentration als Prozess im Zeitablauf.

Das Scientific-Use-File enthält als Merkmale den Umsatz des Berichtsjahres 2000 und des Vorjahres 1999. Für eine dynamische Betrachtung der Konzentration ist dies nicht ausreichend, daher konzentriert sich die folgende Analyse auf die statische Betrachtungsweise.

Problematisch ist die Auswahl der geeigneten Messmethoden. Grundsätzlich existieren *absolute* und *relative* Konzentrationsmaße. Grob formuliert besteht der Unterschied darin, „dass von einer absoluten Konzentration dann gesprochen wird, wenn ein Großteil des gesamten Merkmalsbetrages auf eine kleine *Zahl* von Merkmalsträgern entfällt, von einer relativen Konzentration, wenn ein Großteil des gesamten Merkmalsbetrages auf einen kleinen *Anteil* der Merkmalsträger entfällt“<sup>20)</sup>. Mit den vorliegenden Daten sind Berechnungen mit beiden Arten von Konzentrationsmaßen möglich.

#### Absolute Konzentration

Am einfachsten lässt sich die absolute Konzentration als Anteil der *m* größten Unternehmen am gesamten Merkmalswert berechnen (Konzentrationsrate  $CR_m$ )<sup>21)</sup>. Allerdings blei-

ben dabei sämtliche Informationen unberücksichtigt, die über die anderen (kleineren) Unternehmen bekannt sind. Veränderungen in den Marktanteilen dieser Wettbewerber finden keine Berücksichtigung. Der Vorteil des Maßes liegt in der einfachen Berechenbarkeit, weshalb es sich großer Beliebtheit erfreut.

Ein Alternativmaß, um die Informationen aller am Markt tätigen Unternehmen zu nutzen, ist der Herfindahl-Index. Dieser Index ist gleich der Summe der quadrierten Marktanteile.<sup>22)</sup> Bei absoluter Konzentration erreicht der Index den Wert 1, bei absoluter Gleichverteilung den Wert  $1/\text{Anzahl}$  der Unternehmen. Damit kommt zum Ausdruck, dass die Anzahl der am Markt auftretenden Unternehmen für den Indexwert nicht unerheblich ist: Je größer die Anzahl der Unternehmen ist, desto geringer ist bei einer Gleichverteilung der Indexwert. Durch die Definition werden Unternehmen mit hohen Marktanteilen überproportional gewichtet.

Der größte Vorteil der Umsatzsteuerstatistik liegt in der fast vollständigen Erfassung der Unternehmen. Damit ist der gegenüber den einfacheren Konzentrationsraten aussagefähigere Herfindahl-Index problemlos zu berechnen. Schwierigkeiten bereitet allerdings die aufgrund des Anonymisierungskonzeptes durchgeführte punktuelle Mikroaggregation der drei führenden Unternehmen. Während diese bei Verwendung von Konzentrationsraten keinen Einfluss auf das Ergebnis hat, solange ein *m* mit mindestens der Größe drei gewählt wird, führt sie beim Herfindahl-Index zu einer Unterschätzung. Diese ist umso höher, je höher der Marktanteil des führenden Unternehmens im Verhältnis zu den Marktanteilen der nächstgrößeren Unternehmen ist. Je größer dieser Abstand, desto stärker wird der Marktanteil des führenden Unternehmens durch die Mikroaggregation reduziert. Diese Reduzierung senkt überproportional das Ergebnis – wegen der überproportionalen Gewichtung der Großunternehmen – und damit den Indexwert.

19) Zur Konzentrationsmessung und deren Maßen siehe Vorgrimler, D.: „Wettbewerb aus stagnierenden Märkten“, Band 23 der Schriftenreihe „Spektrum Bundesstatistik“, 2002, S. 75 ff. Im Folgenden wird lediglich die Möglichkeit beschrieben, mit Hilfe der Umsatzsteuerstatistik Konzentrationsmaße zu berechnen. Keine Rolle spielen dagegen weitergehende inhaltliche Fragen der Konzentrationsmessung, wie die Frage nach dem relevanten Markt oder der korrekten Untersuchungseinheit.

20) Bley Müller, J., u. a.: „Statistik für Wirtschaftswissenschaftler“, 12. Auflage, München 2000, S. 191.

21) *m* steht für die wählbare Anzahl der größten Unternehmen, z. B. stellt  $CR_3$  den kumulierten Marktanteil der drei größten Unternehmen am Markt dar.

22) Zur Methode siehe Bley Müller, J., Fußnote 20.

Tabelle 4: Absolute Konzentrationsmaße der Wirtschaftszweige des Abschnitts K

WZ 93 <sup>1)</sup>	Wirtschaftsgliederung	CR <sub>3</sub>			CR <sub>5</sub>			Herfindahl-Index		
		original	anonymisiert	Differenz	original	anonymisiert	Differenz	original	anonymisiert	Differenz
70	Grundstücks- und Wohnungswesen .....	2,59	2,27	-0,32	3,47	3,31	-0,16	6,2	5,7	+0,46
71	Vermietung beweglicher Sachen ohne Bedienungspersonal .....	20,46	19,25	-1,21	26,54	25,49	-1,05	213,9	182,0	+31,82
72	Datenverarbeitung und Datenbanken .....	6,22	6,25	+0,03	8,86	8,89	+0,03	30,1	28,8	+1,36
73	Forschung und Entwicklung .....	38,73	38,77	+0,04	42,98	42,61	-0,37	1104,9	533,2	+571,75
741	Rechts-, Steuer- und Unternehmens- beratung, Markt- und Meinungs- forschung, Beteiligungsgesellschaften ..	13,98	14,10	+0,12	17,33	17,44	+0,11	103,6	84,5	+19,13
742	Architektur- und Ingenieurbüros .....	1,66	1,66	±0,00	2,55	2,55	±0,00	3,8	3,7	+0,02
743	Technische, physikalische und chemische Untersuchungen .....	35,11	35,18	+0,07	48,32	48,36	+0,04	556,4	554,2	+2,20
744	Werbung .....	11,61	11,53	-0,08	15,86	15,79	-0,07	72,8	70,1	+2,75
745	Gewerbsmäßige Vermittlung und Überlassung von Arbeitskräften .....	13,58	13,61	+0,03	18,12	18,16	+0,04	98,8	85,7	+13,15
746	Detekteien und Schutzdienste .....	18,48	18,48	±0,00	22,45	22,46	+0,01	205,6	141,3	+64,30
747	Reinigung von Gebäuden, Inventar und Verkehrsmitteln .....	2,52	2,36	-0,16	3,64	3,36	-0,28	10,0	9,7	+0,27
748	Erbringung von sonstigen Dienst- leistungen für Unternehmen .....	8,36	8,23	-0,13	11,86	11,58	-0,28	48,8	46,6	+2,19

1) Klassifikation der Wirtschaftszweige, Ausgabe 1993.

Beispielhaft wurden aus der Umsatzsteuerstatistik sowohl aus den Original- als auch aus den anonymisierten Daten die beschriebenen absoluten Konzentrationsmaße für die Wirtschaftszweige des Abschnitts K der WZ 93 berechnet (siehe Tabelle 4). Die Konzentrationsraten wurden hierbei für  $m = 3$  und  $m = 5$  ermittelt. Während die Konzentrationsraten nur minimale bis keine Veränderungen aufgrund der Anonymisierung aufweisen, wird der Herfindahl-Index wie erwartet bei steigender Konzentration unterschätzt. Die Korrelation zwischen der Unterschätzung des Herfindahl-Index und der Höhe der Konzentrationsraten liegt über alle Wirtschaftszweige berechnet bei 0,75. Werden daher Konzentrationsuntersuchungen mit absoluten Konzentrationsmaßen mit Hilfe der anonymisierten Umsatzsteuerstatistik durchgeführt, so sollten neben dem theoretisch aussagefähigeren Herfindahl-Index ebenfalls Konzentrationsraten berechnet werden, um das Ausmaß der Unterschätzung des Herfindahl-Index bei Branchen mit hoher Konzentration einschätzen zu können.

### Relative Konzentration

Mit der relativen Konzentration, die im Folgenden mit Hilfe des Gini-Koeffizienten berechnet wird, kann die Verteilung innerhalb einer Gruppe untersucht werden. Der Gini-Koeffizient ist ein Maß zur Messung der Krümmung der so genannten Lorenz-Kurve. Bei der Lorenz-Kurve handelt es sich um eine Darstellungsform der relativen Konzentration, bei der die kumulierten relativen Häufigkeiten des Merkmals „Unternehmen“ und die relativen Anteile am Gesamtumsatz in Beziehung gesetzt werden.<sup>23)</sup> Werden die Unternehmen dabei nach der Größe des Umsatzes (beginnend mit dem kleinsten) kumuliert, so erscheint die Lorenz-Kurve umso stärker nach unten gewölbt, je größer die relative Konzentration ist. Bei einer absoluten Gleichverteilung des Umsatzes entsteht eine Gerade (45°-Linie).

Im Gegensatz zu den Maßen der absoluten Konzentration sind keine Vorabausagen darüber möglich, wie sich die Anonymisierung auf den Gini-Koeffizienten auswirkt. Veränderungen bei Großunternehmen können durch Veränderungen bei kleineren Unternehmen wieder ausgeglichen werden, da alle Unternehmen mit dem gleichen Gewicht in den Koeffizienten eingehen. Aussagen zu den Auswirkungen können daher nur anhand von Beispielsrechnungen getroffen werden.

Wie Tabelle 5 zeigt, bewirken die Anonymisierungsmaßnahmen beim Gini-Koeffizienten nur minimale bzw. keine Ver-

Tabelle 5: Gini-Koeffizienten für die Wirtschaftszweige des Abschnitts K „Grundstücks- und Wohnungswesen, Vermietung beweglicher Sachen, Erbringung von wirtschaftlichen Dienstleistungen, anderweitig nicht genannt“

WZ 93 <sup>1)</sup>	Wirtschaftsgliederung	Gini-Koeffizient original	Gini-Koeffizient anonymisiert
70	Grundstücks- und Wohnungswesen .....	0,8500	0,8499
71	Vermietung beweglicher Sachen ohne Bedienungspersonal .....	0,9181	0,9168
72	Datenverarbeitung und Datenbanken .....	0,8739	0,8739
73	Forschung und Entwicklung .....	0,9142	0,9142
741	Rechts-, Steuer- und Unternehmensberatung, Markt- und Meinungsforschung, Beteiligungsgesellschaften .....	0,8663	0,8665
742	Architektur- und Ingenieurbüros .....	0,7444	0,7443
743	Technische, physikalische und chemische Untersuchungen .....	0,9308	0,9308
744	Werbung .....	0,8747	0,8745
745	Gewerbsmäßige Vermittlung und Überlassung von Arbeitskräften .....	0,7576	0,7577
746	Detekteien und Schutzdienste .....	0,8661	0,8661
747	Reinigung von Gebäuden, Inventar und Verkehrsmitteln .....	0,7715	0,7714
748	Erbringung von sonstigen Dienstleistungen für Unternehmen ..	0,8856	0,8852

1) Klassifikation der Wirtschaftszweige, Ausgabe 1993.

23) Zur Lorenz-Kurve und dem Gini-Koeffizienten siehe Henze, A.: „Marktforschung“, Stuttgart 1994, S. 322 ff.

änderungen. Dies gilt nicht nur für die dargestellten Branchen, sondern auch für die weiteren im Scientific-Use-File enthaltenen Wirtschaftszweige. Untersuchungen der relativen Konzentration werden demnach durch die Anonymisierung nicht beeinträchtigt und können daher ohne Einschränkungen durchgeführt werden.

### 3.3 Weitere Analysemöglichkeiten

Neben den unmittelbar aus den Daten der Finanzverwaltung ableitbaren Erkenntnissen lässt sich die Umsatzsteuerstatistik auch als Basis für weitere wirtschaftsstatistische Analysen nutzen.

Nach § 15 Abs. 2 Umsatzsteuergesetz (UStG) sind steuerfreie Lieferungen und Leistungen vom Vorsteuerabzug ausgeschlossen. Dieser grundsätzliche Ausschluss gilt jedoch nicht bei Lieferungen in Mitgliedstaaten der EU sowie Ausfuhren in Drittländer. Eine Steuerfreiheit mit Vorsteuerabzug ist darüber hinaus beispielsweise für Umsätze der Seeschifffahrt und der Luftfahrt, Goldlieferungen an Zentralbanken und Lieferungen an Vertragsparteien des Nordatlantikvertrages vorgesehen. Diese Lieferungen sind nicht quantifizierbar, dürften jedoch insgesamt nicht von größerer Bedeutung sein, sodass anhand des Merkmals der steuerfreien Lieferungen und Leistungen mit Vorsteuerabzug die Exporttätigkeit der Unternehmen beschrieben werden kann.

Tabelle 6 zeigt die aus den steuerfreien Lieferungen und Leistungen mit Vorsteuerabzug und den Lieferungen und Leistungen (Umsatz) berechneten Exportquoten der Unternehmen in den im anonymisierten Datenfile nachgewiese-

Tabelle 6: Exportquoten<sup>1)</sup> 2000 der Unternehmen in den Wirtschaftszweigen des Abschnittes K  
Prozent

WZ 93 <sup>2)</sup>	Wirtschaftsgliederung	Exportquote <sup>1)</sup>	
		original	anonymisiert
70	Grundstücks- und Wohnungswesen .....	3,2	3,2
71	Vermietung beweglicher Sachen ohne Bedienungspersonal .....	7,7	7,8
72	Datenverarbeitung und Datenbanken .....	5,0	5,0
73	Forschung und Entwicklung .....	41,3	38,6
741	Rechts-, Steuer- und Unternehmensberatung, Markt- und Meinungsforschung, Beteiligungsgesellschaften .....	13,8	14,0
742	Architektur- und Ingenieurbüros ...	3,4	3,4
743	Technische, physikalische und chemische Untersuchung .....	3,7	3,7
744	Werbung .....	1,9	1,9
745	Gewerbsmäßige Vermittlung und Überlassung von Arbeitskräften ..	0,4	0,4
746	Detekteien und Schutzdienste .....	2,4	2,4
747	Reinigung von Gebäuden, Inventar und Verkehrsmitteln .....	0,9	0,9
748	Erbringung von sonstigen Dienstleistungen überwiegend für Unternehmen .....	10,0	9,9
	Zusammen ...	7,5	7,5

1) Verhältnis des steuerfreien Umsatzes mit Vorsteuerabzug zum Gesamtumsatz in Prozent. - 2) Klassifikation der Wirtschaftszweige, Ausgabe 1993.

nen Wirtschaftszweigen des Wirtschaftsabschnitts K „Grundstücks- und Wohnungswesen, Vermietung beweglicher Sachen, Erbringung von Dienstleistungen, überwiegend für Unternehmen“. Die größte Umsatzbedeutung haben die Exporte in den Bereichen 73 „Forschung und Entwicklung“ und bei den Unternehmen der „Rechts-, Steuer- und Unternehmensberatung, Markt- und Meinungsforschung, Beteiligungsgesellschaften“ (WZ 74.1).

Weitere interessante Analysemöglichkeiten bietet die Auswertung der Exportquoten der Unternehmen nach Umsatzgrößenklassen. Tabelle 7 zeigt die Exportquoten der Unternehmen nach Umsatzgrößenklassen über alle Wirtschaftszweige. Dabei wird erwartungsgemäß eine deutlich stärkere Exportorientierung der Großunternehmen deutlich. Bei den Unternehmen mit mehr als 250 Mill. Euro Umsatz wird nahezu jeder vierte Euro mit dem Ausland umgesetzt. Aber auch die mittelständischen Unternehmen der Umsatzgrößenklasse 25 bis unter 50 Mill. Euro erwirtschaften 17% ihrer Umsätze mit dem Ausland.

Tabelle 7: Exportquoten<sup>1)</sup> 2000 der Unternehmen nach Umsatzgrößenklassen  
Prozent

Größenklassen der Lieferungen und Leistungen von ... bis unter ... EUR	Exportquote <sup>1)</sup>	
	original	anonymisiert
unter 1 Mill. ....	1,8	1,8
1 Mill. – 5 Mill. ....	5,5	5,5
5 Mill. – 25 Mill. ....	11,3	11,4
25 Mill. – 50 Mill. ....	16,1	16,1
50 Mill. – 100 Mill. ....	17,9	17,9
100 Mill. – 250 Mill. ....	20,1	20,3
250 Mill. und mehr .....	23,4	23,3
Insgesamt ...	16,1	16,1

1) Verhältnis des steuerfreien Umsatzes mit Vorsteuerabzug zum Gesamtumsatz in Prozent.

Neben diesen Auswertungen können die Exportquoten auch mit allen anderen qualitativen Merkmalen des Mikrodatenfiles kombiniert werden (z. B. Rechtsform der Unternehmen oder Regionalgliederung). Zudem werden die steuerfreien Umsätze mit Vorsteuerabzug für innergemeinschaftliche Lieferungen an Abnehmer mit Umsatzsteuer-Identifikationsnummer (§ 4 Nr. 1b UStG) gesondert nachgewiesen, sodass sich auch die EU-Exporte gezielt untersuchen lassen.

Umfassende Importquoten lassen sich mit der Umsatzsteuerstatistik hingegen nicht ermitteln. Während Importe aus EU-Mitgliedstaaten als innergemeinschaftliche Erwerbe im Rahmen der Umsatzsteuer-Voranmeldungen angegeben werden, wird für Importe aus Drittländern die so genannte Einfuhrumsatzsteuer von der Zollverwaltung erhoben. Daher lassen sich anhand des Datenfiles lediglich „EU-Importe“ näher untersuchen.

Ein weiteres Forschungsfeld ist die Analyse der Auswirkungen von Steuerrechtsänderungen im Bereich der Umsatzsteuer. Hierbei kann zum Beispiel untersucht werden, in welchem quantitativen Ausmaß sich eine Abschaffung des ermäßigten Umsatzsteuersatzes oder eine generelle Erhö-

hung desselben auswirken. Wie aktuelle Untersuchungen zeigen, spielen solche steuerpolitischen Themen in der öffentlichen Diskussion immer wieder eine große Rolle.<sup>24)</sup>

Die Frage, welche Branchen am meisten von einem ermäßigten Umsatzsteuersatz profitieren, ist ebenfalls eine steuerpolitische Fragestellung, die mit den Daten der Umsatzsteuerstatistik beantwortet werden kann. Tabelle 8 zeigt die zehn im anonymisierten Datenfile gesondert nachgewiesenen Branchen, welche die höchsten Anteile ihres Umsatzes mit Produkten erwirtschaften, die dem ermäßigten Steuersatz von 7% unterliegen. Die Anteile wurden wiederum einmal mit den anonymisierten Daten und einmal mit den Originaldaten berechnet. Die Ergebnisse unterscheiden sich dabei nur unerheblich.

Tabelle 8: Die zehn Branchen mit den höchsten Anteilen der mit dem ermäßigten Umsatzsteuersatz von 7% besteuerten Umsätze 2000  
Prozent

WZ 93 <sup>1)</sup>	Wirtschaftsgliederung	Anteil	
		original	anonymisiert
41 DA	Wasserversorgung .....	59,5	59,5
	Ernährungsgewerbe und Tabakverarbeitung .....	57,5	57,7
522	Facheinzelhandel mit Nahrungs- mitteln, Getränken und Tabak- waren (in Verkaufsräumen) .....	55,3	55,2
A,B	Land- und Forstwirtschaft, Fischerei und Fischzucht .....	49,4	49,4
521	Einzelhandel mit Waren verschiedener Art (in Verkaufsräumen) .....	46,2	45,0
22	Verlagsgewerbe, Druckgewerbe, Vervielfältigung von bespielten Ton-, Bild- und Datenträgern .....	26,3	26,0
91	Interessenvertretung sowie kirchliche und sonstige religiöse Vereinigungen (ohne Sozialwesen und Sport) .....	21,4	21,4
55B	Gaststättengewerbe und Kantinen .	20,1	20,1
92	Kultur, Sport und Unterhaltung ....	19,6	19,9
51	Handelsvermittlung und Groß- handel (ohne Kfz) .....	16,6	16,6

1) Klassifikation der Wirtschaftszweige, Ausgabe 1993.

Durch das aufgenommene Merkmal „Neugründung“ sind auch Analysen zum Gründungsgeschehen mit dem Scientific-Use-File durchführbar. Die Möglichkeiten und Grenzen, dies mit den Daten der Umsatzsteuerstatistik zu tun, werden in Gräß/Zwick beschrieben.<sup>25)</sup>


## 4 Fazit

Nach der Lohn- und Einkommensteuerstatistik 1998 kann mit den Daten zur Umsatzsteuerstatistik 2000 der Wissenschaft ein zweiter Datenbestand der Steuerstatistiken in Form eines standardisierten Scientific-Use-File angeboten werden. Die zur Geheimhaltung getroffenen Maßnahmen wurden auf das Notwendigste beschränkt, sodass mög-

lichst viel vom Informationsgehalt der Originaldaten im Scientific-Use-File erhalten wurde. Das Potenzial der Umsatzsteuerstatistik als Quelle wirtschaftsstatistischer Analysen kann damit von der externen Wissenschaft mit wenigen Einschränkungen erschlossen werden.

Auch wenn die Anonymisierung auf das Notwendigste reduziert wurde, werden einige Analysen für einen Wissenschaftler aufgrund der Einschränkungen nicht möglich sein. In solchen Fällen eröffnet die deutsche amtliche Statistik durch die Einrichtung der Forschungsdatenzentren alternative Zugangswege, mit Hilfe derer auch diese Analysen möglich werden. Zu denken wäre dabei zum Beispiel an tiefergehende Regionalanalysen, die mit Hilfe von Sonderauswertungen in den Fachabteilungen durchgeführt werden könnten.

Das Scientific-Use-File der Umsatzsteuerstatistik 2000 ist ein weiterer Schritt hin zu einem umfassenden Datenangebot an die Wissenschaft aus dem Bereich der Steuerstatistik. Speziell bei der Umsatzsteuerstatistik ergeben sich darüber hinaus zahlreiche neue Analysemöglichkeiten, wenn von einer reinen Querschnitts- zu einer Längsschnittsbetrachtung übergegangen wird. So können Fragen der Unternehmensdemographie (z. B. die Entwicklung von neu gegründeten Unternehmen) oder der dynamischen Konzentration mit einem Scientific-Use-File der Umsatzsteuerstatistik analysiert werden, wenn dieses als Panel über mehrere Jahre vorliegt.

Das Scientific-Use-File ist für 65,- Euro 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. 

24) Siehe z. B. Zentrum für Europäische Wirtschaftsforschung (ZEW): „Ermäßigter Umsatzsteuersatz als Instrument der Verteilungspolitik ungeeignet“, Pressemitteilung, 5. Januar 2005.

25) Gräß, C./Zwick M.: „Die Umsatzsteuerstatistik“ in Fritsch, M./Grotz, R. (Hrsg.): „Das Gründungsgeschehen in Deutschland – Darstellung und Vergleich der Datenquellen“, Heidelberg 2002, S. 129 ff.

WZ-Klassifizierung des Scientific-Use-File der Umsatzsteuerstatistik 2000

WZ 93 <sup>1)</sup>	Bezeichnung
A, B .....	Land- und Forstwirtschaft, Fischerei und Fischzucht
C .....	Bergbau und Gewinnung von Steinen und Erden
DA .....	Ernährungsgewerbe und Tabakverarbeitung
DB .....	Textil- und Bekleidungsgewerbe
19 .....	Ledergewerbe
20 .....	Holzgewerbe (ohne Herstellung von Möbeln)
21 .....	Papiergewerbe
22 .....	Verlagsgewerbe, Druckgewerbe, Vervielfältigung von bespielten Ton-, Bild- und Datenträgern
23 .....	Kokerei, Mineralölverarbeitung, Herstellung und Verarbeitung von Spalt- und Brutstoffen
24 .....	Chemische Industrie
25 .....	Herstellung von Gummi- und Kunststoffwaren
26 .....	Glasgewerbe, Keramik, Verarbeitung von Steinen und Erden
27 .....	Metallerzeugung und -bearbeitung
28 .....	Herstellung von Metallerzeugnissen
29 .....	Maschinenbau
30 .....	Herstellung von Büromaschinen, Datenverarbeitungsgeräten und -einrichtungen
31 .....	Herstellung von Geräten der Elektrizitätserzeugung, -verteilung u. Ä.
32 .....	Rundfunk-, Fernseh- und Nachrichtentechnik
33 .....	Medizin-, Mess-, Steuer- und Regelungstechnik, Optik
34 .....	Herstellung von Kraftwagen und Kraftwagenteilen
35 .....	Sonstiger Fahrzeugbau
36 .....	Herstellung von Möbeln, Schmuck, Musikinstrumenten, Sportgeräten, Spielwaren und sonstigen Erzeugnissen
37 .....	Recycling
40 .....	Energieversorgung
41 .....	Wasserversorgung
45.A .....	Bauhauptgewerbe (45.1, 45.2)
45.B .....	Baubausgewerbe (45.3 – 45.5)
50 .....	Kraftfahrzeughandel; Instandhaltung und Reparatur von Kraftfahrzeugen; Tankstellen
51 .....	Handelsvermittlung und Großhandel (ohne Handel mit Kraftfahrzeugen)
52.1 .....	Einzelhandel mit Waren verschiedener Art (in Verkaufsräumen)
52.2 .....	Facheinzelhandel mit Nahrungsmitteln, Getränken und Tabakwaren (in Verkaufsräumen)
52.3 .....	Apotheken; Facheinzelhandel mit medizinischen, orthopädischen und kosmetischen Artikeln (in Verkaufsräumen)
52.4 .....	Sonstiger Facheinzelhandel (in Verkaufsräumen)
52.5 .....	Einzelhandel mit Antiquitäten und Gebrauchsgütern (in Verkaufsräumen)
52.6 .....	Einzelhandel (nicht in Verkaufsräumen)
52.7 .....	Reparatur von Gebrauchsgütern
55.A .....	Beherbergungsgewerbe (55.1, 55.2)
55.B .....	Gaststättengewerbe und Kantinen (55.3 – 55.5)
60 .....	Landverkehr; Transport in Rohrfernleitungen
61 .....	Schifffahrt
62 .....	Luftfahrt
63.1 .....	Frachttumschlag und Lagerei
63.2 .....	Sonstige Hilfs- und Nebentätigkeiten für den Verkehr
63.3 .....	Reisebüros und Reiseveranstalter
63.4 .....	Spedition, sonstige Verkehrsvermittlung
64 .....	Nachrichtenübermittlung
J .....	Kredit- und Versicherungsgewerbe
70 .....	Grundstücks- und Wohnungswesen
71 .....	Vermietung beweglicher Sachen ohne Bedienungspersonal
72 .....	Datenverarbeitung und Datenbanken
73 .....	Forschung und Entwicklung
74.1 .....	Rechts-, Steuer- und Unternehmensberatung, Markt- und Meinungsforschung, Beteiligungsgesellschaften
74.2 .....	Architektur- und Ingenieurbüros
74.3 .....	Technische, physikalische und chemische Untersuchung
74.4 .....	Werbung
74.5 .....	Gewerbsmäßige Vermittlung und Überlassung von Arbeitskräften
74.6 .....	Detekteien und Schutzdienste
74.7 .....	Reinigung von Gebäuden, Inventar und Verkehrsmitteln
74.8 .....	Erbringung von sonstigen Dienstleistungen überwiegend für Unternehmen
L, M, N .....	Öffentliche Verwaltung, Verteidigung, Sozialversicherung, Erziehung und Unterricht, Gesundheits-, Veterinär- und Sozialwesen
90 .....	Abwasser- und Abfallbeseitigung und sonstige Entsorgung
91 .....	Interessenvertretungen sowie kirchliche und sonstige religiöse Vereinigungen (ohne Sozialwesen und Sport)
92 .....	Kultur, Sport und Unterhaltung
93.01 .....	Wäscherei und chemische Reinigung
93.02 .....	Friseurgewerbe und Kosmetiksalons
93.03 .....	Bestattungswesen
93.04 .....	Bäder, Saunas, Solarien u. Ä.
93.05 .....	Erbringung von Dienstleistungen a. n. g.

1) Klassifikation der Wirtschaftszweige, Ausgabe 1993.