

FDZ-Arbeitspapier  
Nr. 13

Daniel Vorgrimler



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

Anonymisierte Daten der  
amtlichen Steuerstatistik

2006

FDZ-Arbeitspapier  
Nr. 13

Daniel Vorgrimler



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

Anonymisierte Daten der  
amtlichen Steuerstatistik

2006

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

**Fachliche Informationen**

zu dieser Veröffentlichung:

Statistisches Bundesamt  
Forschungsdatenzentrum  
Tel.: 06 11 / 75 42 20  
Fax: 06 11 / 72 40 00  
forschungsdatenzentrum@destatis.de

Erscheinungsfolge: unregelmäßig  
Erschienen im September 2006

**Informationen zum Datenangebot:**

Statistisches Bundesamt  
Forschungsdatenzentrum  
Tel.: 06 11 / 75 42 20  
Fax: 06 11 / 72 40 00  
forschungsdatenzentrum@destatis.de

Forschungsdatenzentrum der  
Statistischen Landesämter  
– Geschäftsstelle –  
Tel.: 0211 / 9449 41 47  
Fax: 0211 / 9449 40 77  
forschungsdatenzentrum@lds.nrw.de

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

© Statistisches Bundesamt, Wiesbaden 2006  
(im Auftrag der Herausbergemeinschaft)

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.

# Anonymisierte Daten der amtlichen Steuerstatistik

*Daniel Vorgrimler<sup>\*)</sup>*

## 1 Einführung

Die auf Mikrodaten basierende wissenschaftliche Analyse von Steuerrechtsänderungen gewinnt zunehmend an Bedeutung. Als Grundlage für diese Analysen dienen Daten, welche die Entstehung der Steuerschuld möglichst genau abbilden. Die amtliche Statistik kann mit den Mikrodaten der verschiedenen Steuerstatistiken diese Anforderungen erfüllen. Darüber hinaus sind mit diesen Mikrodaten Analysen möglich, die weit über den originären Bereich der Steuer- und Steuerlastsimulation hinausgehen.

Um die Mikrodaten der wissenschaftlichen Analyse bereitzustellen, hat der Gesetzgeber als einen Weg der Datennutzung mit dem § 16 Abs. 6 BStatG der Wissenschaft einen privilegierten Datenzugang eingeräumt. Mikrodaten, die den Geheimhaltungsvorschriften des § 16 Abs. 6 BStatG genügen, werden aufgrund dieses Wissenschaftsprivilegs Scientific-Use-File genannt. Scientific-Use-Files sind darüber hinaus Daten, die außerhalb der amtlichen Statistik von Mitarbeitern wissenschaftlicher Einrichtungen genutzt werden können. Im folgenden Beitrag werden in den Abschnitten zwei und drei die zum Stand Oktober 2005 angebotenen Scientific-Use-Files der amtlichen Steuerstatistik mit ihren Möglichkeiten und Grenzen vorgestellt. Darüber hinaus gibt Abschnitt 4 einen Ausblick auf Entwicklungen beim Datenangebot der amtlichen Steuerstatistik. Der Schwerpunkt liegt hierbei bei der Entwicklung und Erschließung neuer Datenangebote. Ein Fazit rundet den Beitrag ab.

## 2 Faktisch anonymisierte Lohn- und Einkommensteuerstatistik 1998 (FAST 98)

### 2.1 Datengrundlage<sup>1</sup>

Die Lohn- und Einkommensteuerstatistik ist eine dezentrale dreijährliche Sekundärstatistik. Die Finanzverwaltungen liefern die jeweiligen Angaben der Steuerpflichtigen zu vorgegebenen Terminen an die Statistischen Landesämter. Diese generieren die jeweiligen Landesergebnisse und übermitteln diese an das Statistische Bundesamt. Das Statistische Bundesamt führt dann die Länderergebnisse im nächsten Schritt zum Bundesergebnis zusammen. Durch die Novellierung des Gesetzes über Steuerstatistiken (StStatG) im

---

<sup>\*)</sup> Dr. Daniel Vorgrimler, Statistisches Bundesamt, Wiesbaden.

<sup>1</sup> Ausführlich zur Lohn- und Einkommensteuerstatistik vgl. Rosinus, W. (2000)

Rahmen des Jahressteuergesetzes 1996 werden neben den Tabellendaten zusätzlich die Einzelangaben (Mikrodaten) von den Statistischen Landesämtern an das Statistische Bundesamt übermittelt.

Als Sekundärstatistik ist die Lohn- und Einkommensteuerstatistik abhängig von den Einkommensteuer-  
veranlagungen, die von den Finanzverwaltungen durchgeführt werden. Aufgrund der den Steuerpflichtigen  
zugestandenen Fristen zur Einreichung ihrer Einkommensteuererklärung vergehen 2 3/4 Jahre bis die letz-  
ten Daten den jeweiligen Statistischen Landesämtern zur Verfügung stehen. Die Dreijährlichkeit der Statis-  
tik und die Fristen zur Einkommensteueranmeldung haben zur Folge, dass erst im vierten Jahr nach Ende  
des betreffenden Veranlagungsjahres Ergebnisse vorliegen und diese zum Teil bis in das siebte Jahr die  
aktuellsten bleiben.

Aufgrund ihrer Datenvielfalt bietet die Lohn- und Einkommensteuerstatistik vielfältige Analysemöglichkei-  
ten. Neben steuerlichen Betrachtungen können auch Untersuchungen über die Einkommensverteilung  
durchgeführt werden. Besonders die Bezieher hoher und höchster Einkommen sind in keiner anderen sta-  
tistischen Quelle so genau erfasst wie in der Lohn- und Einkommensteuerstatistik. Bei Analysen muss al-  
lerdings beachtet werden, dass die Einkommensbegriffe der Lohn- und Einkommensteuerstatistik auf dem  
Steuerrecht basieren. Daher sind die Merkmale nicht ohne weiteres mit denen aus der Volkswirtschaftli-  
chen Gesamtrechnung (VGR) vergleichbar. Dem Einkommensbegriff der VGR am nächsten kommt der Beg-  
riff „Gesamtbetrag der Einkünfte“ (GdE). Aber auch dieser berücksichtigt z. B. nur zum Teil staatliche Um-  
verteilungen, welche das tatsächlich verfügbare Einkommen der Haushalte ebenfalls beeinflussen.

Bei den Überschusseinkünften<sup>2</sup> kann deren Entstehung betrachtet werden (z. B. Bruttolohn minus die  
Werbungskosten). Dies ist bei den Gewinneinkünften<sup>3</sup> nicht möglich, da keine Angaben über die Be-  
triebseinnahmen und -ausgaben in den Daten enthalten sind.

Ein Datensatz repräsentiert einen Steuerpflichtigen. Bei einer gemeinsamen Veranlagung von Ehepaaren  
im Splittingfall besteht ein Steuerpflichtiger aus zwei Personen bzw. zwei Steuerfällen. Daher umfassen  
die knapp 30 Mill. Einzeldatensätze Angaben von über 42 Mill. Steuerfällen. Bis zum Merkmal „Summe  
der Einkünfte“ werden dabei die jeweiligen Merkmale für die Ehepartner getrennt ausgewiesen. Im weite-  
ren Besteuerungsverlauf ist dies nicht mehr sinnvoll. Als Folge der Unterscheidung zwischen Steuerpflich-  
tigen und Steuerfällen ist die steuerliche Einkommensverteilung basierend auf der Verteilung des „Ge-  
samtbetrags der Einkünfte“ keine Verteilung der Individualeinkommen. Sie bildet auch nicht exakt die Ver-  
teilung der Haushaltseinkommen ab, da innerhalb eines Haushaltes mehrere Steuerpflichtige leben kön-  
nen. Dennoch wird in Analysen i. d. R. der Steuerpflichtige als Approximation des Haushaltes verwendet.<sup>4</sup>

---

<sup>2</sup> Zu den Überschusseinkünften zählen Einkünfte aus nichtselbständiger Arbeit, aus Vermietung und Verpachtung, aus Kapitalvermögen sowie aus sonstigen Einkünften.

<sup>3</sup> Zu den Gewinneinkünften zählen Einkünfte aus Land- und Forstwirtschaft, aus Gewerbebetrieb sowie aus selbständiger Arbeit.

<sup>4</sup> Z.B. vgl. Bach, S., Haan, P., Rudolph, H.-J., Steiner, V. (2004) sowie Rosinus, W. (2000)

Die Datensätze weisen neben den quantitativen Merkmalen des Besteuerungsprozesses auch sozioökonomische Merkmale aus, die eine gezielte Analyse einzelner gesellschaftlicher Gruppen ermöglichen. Zu diesen Merkmalen zählen u. a. die regionale Gliederung, das Alter und bei Gewerbetreibenden der Wirtschaftszweig.

## 2.2 Anonymisierungskonzeption

Faktisch anonyme Daten sollten nur dann von den Statistischen Ämtern als Scientific-Use-File der Wissenschaft zur Verfügung gestellt werden, wenn sie ausreichende wissenschaftliche Analysemöglichkeiten bieten. Da eine Anonymisierung von Merkmalsträgern immer eine Reduktion von Information impliziert, sollte daher die Anonymisierung auf das Notwendigste beschränkt sein. Um dies zu erreichen sind bei FAST 98 die Merkmalsträger in Abhängigkeit zu ihrem Reidentifikationsrisiko anonymisiert. Diejenigen mit einem geringeren Reidentifikationsrisiko sind entsprechend schwächer anonymisiert als diejenigen mit erhöhten Risiken. Dabei wurde angenommen, dass das Risiko der Reidentifikation mit der Einkommenshöhe zunimmt. Auf Grundlage dieser Annahme wurden die Merkmalsträger in fünf verschiedene Einkommensbereiche eingeteilt und erhielten so einen Indikator für ihr Risiko. Innerhalb dieser Bereiche, die im Folgenden Anonymisierungsbereiche genannt werden, wurden speziell auf das Risiko abgestimmte Anonymisierungsmaßnahmen durchgeführt.<sup>5</sup> Als Datenbasis für die Anonymisierung der Lohn- und Einkommensteuerstatistik 1998 diente die 10%-Stichprobe, zu deren Ziehung das Statistische Bundesamt aufgrund des § 7 Abs. 4 StStatG verpflichtet ist.

Tabelle 1 gibt die Bedeutung der Anonymisierungsbereiche nach den Kriterien Steuerpflichtige, Gesamtbeitrag der Einkünfte und festgesetzte Einkommensteuer wieder. Sie zeigt, dass dem Anonymisierungsbereich 1 die größte Bedeutung zukommt.

**Tabelle 1: Anteile der Anonymisierungsbereiche**

Bereich	Steuerpflichtige		GdE		Einkommensteuer	
	Anteil	kumuliert	Anteil	kumuliert	Anteil	kumuliert
1	92,2	92,2	68,8	68,8	51,7	51,7
2	6,6	98,8	17,1	85,9	22,2	73,9
3	1,13	99,93	8,6	94,5	16,2	90,1
4	0,05	99,98	3,2	97,7	6,2	96,3
5	0,02	100	2,3	100	3,7	100

Quelle: Merz, J., Vorgrimler, D., Zwick, M. (2004)

Neben der auf die Einkommenshöhe abgestimmten Anonymisierung wurden Maßnahmen ergriffen, mit denen alle Merkmalsträger anonymisiert wurden (allgemeine Anonymisierung). Übersicht 1 gibt über diese Maßnahmen Auskunft.

<sup>5</sup> Ausführlich zur Anonymisierungskonzeption und zur Einteilung der Bereiche vgl. Merz, J., Vorgrimler, D., Zwick, M (2004)

## Übersicht 1: Allgemeine Anonymisierungsmaßnahmen

Merkmal	Verbliebene Ausprägungen
Religion (getrennt nach Männer und Frauen)	1 = evangelisch 2 = katholisch 3 = sonstige 4 = konfessionslos
Veranlagungsart	1 = Grundtabelle 2 = Splittingtabelle
Alter (getrennt nach Männer und Frauen)	Einführung einer Unter- (15 Jahre) und Obergrenze (70 Jahre). Ober- bzw. unterhalb der Grenzen wurde das Alter als Durchschnitt derjenigen, die ober- bzw. unterhalb der Grenzen liegen, angegeben.
Anzahl der Kinder	Anzahl der Kinder bis vier; Alter der ersten drei Kinder
GKZ	Ausgewiesen auf Einstellerebene

Quelle: Merz, J., Vorgrimler, D., Zwick, M. (2004)

In den fünf Anonymisierungsbereichen wurden Merkmale unterschiedlich vergrößert oder gestrichen (vgl. Übersicht 2). Hierzu wurden die stetigen Merkmale – wie z. B. die Einkommensmerkmale – nach ihrer Bedeutung in drei Kategorien eingeteilt. In der Ersten sind die wichtigsten Merkmale enthalten, die auch bei den Merkmalsträgern mit den höchsten Einkommen noch ausgewiesen werden (z. B. festzusetzende Einkommensteuer, Gesamtbetrag der Einkünfte). Die zweite Kategorie enthält Merkmale, die nur bei den höchsten Einkommen behandelt werden (z. B. die sieben Einkunftsarten, die Sonderausgaben und die außergewöhnlichen Belastungen), während die Merkmale der dritten Kategorie als erstes eingeschränkt werden.

## Übersicht 2: Spezielle Anonymisierungsmaßnahmen in den Einkommensbereichen

Merkmal	Bereiche <sup>1</sup>				
	1	2	3	4	5
Religion	4 Ausprägungen	4 Ausprägungen	k. A.	k. A.	k. A.
Kinder	Anzahl bis 4 Alter der ersten 3 Kinder	Anzahl bis 4 Alter als Dummy	Anzahl bis 4 Alter als Dummy	Anzahl bis 4	Ja/nein
Alter	Ja mit 15 / 70 Grenze	Klasse mit 5 Jahren	Klasse mit 10 Jahren	Klasse mit 10 Jahren	Klasse mit 10 Jahren
Region	Bundesland	Bundesland	West/Ost	West/Ost	West/Ost
GKZ	1-Steller	1-Steller	1-Steller	1-Steller	k. A.
Stetige Merkmale	1	Ja	Ja	Ja	Ja
	2	Ja	Ja	Ja	Ja, männlich/weiblich summiert
	3	Ja	Ja	Ja	Dummy
				Dummy	Nein

<sup>1</sup> Bei positiven Einkommen: 1 = 0 bis zu einem GdE von 64.106€; 2 = 64.107 – 137.532 €; 3 = 137.533 – 970.202 €; 4 = 970.203 € bis zum 1.000 höchsten GdE; 5 = die 1.000 höchsten GdE + Abgeordnete.

Bei negativen Einkommen: 1 = 0 bis zu einem negativen Einkommen von 102.258 €; 3 = 102.259 bis zu einem negativen Einkommen von 511.292 €; 5 = bei einem negativen Einkommen von über 511.292 €.

Informationen, die aufgrund der Anonymisierung entweder nur vergrößert, verfälscht oder überhaupt nicht mehr in den Zieldaten enthalten sind, weisen für einen Datenangreifer einen geringeren Wert auf als die Originalinformationen. Anonymisierung wirkt sich daher nicht nur auf die Kosten eines Datenangreifers aus, sondern sein Nutzen wird ebenfalls negativ beeinflusst.<sup>6</sup> Bei der Lohn- und Einkommensteuerstatistik gilt dieser Aspekt besonders bei den stetigen Merkmalen. Diese sind evtl. schwierig als Überschneidungsmerkmale einsetzbar, wodurch eine Veränderung ihrer Werte keinen zusätzlichen Schutz der Merkmalsträger darstellen würde, jedoch dürften die stetigen Merkmale einem Datenangreifer den höchsten Nutzen stiften. Werden daher stetige Merkmale aus den Daten gelöscht oder vergrößert, so hat dies vor allem auf die Nutzenseite des Datenangriffs eine Wirkung. Dies ist ein wesentlicher Aspekt zur Erreichung der faktischen Anonymität, bei der die Unverhältnismäßigkeit eines Datenangriffs mitberücksichtigt wird.

Die an der Diskussion über das Anonymisierungskonzept beteiligten Wissenschaftler bevorzugen den Erhalt der stetigen Merkmale gegenüber den sozioökonomisch diskreten Merkmalen, wie z. B. das Alter oder die Religion. Dementsprechend sieht das Anonymisierungskonzept vor, zunächst die diskreten Merkmale zu vergrößern bzw. zu löschen, bevor die stetigen Merkmale zur Anonymisierung herangezogen werden (vgl. Übersicht 2). So ist z. B. das Merkmal „Religion“ nur in den ersten beiden Bereichen mit vier Ausprägungen vertreten.

<sup>6</sup> Vgl. hierzu Höhne J., Sturm R., Vorgrimler D. (2003) sowie Lenz, R., Sturm, R., Vorgrimler, D. (2004)



Aufgrund des stärkeren Eingriffs in die sozioökonomischen Merkmale konnten die stetigen bis einschließlich des dritten Bereichs unverändert in den Daten verbleiben. Im vierten Bereich sind die Merkmale der dritten Kategorie noch als Dummy-Variable enthalten, während sie im fünften gelöscht sind. Die Merkmale der zweiten Kategorie werden im vierten Bereich weiterhin ausgewiesen, allerdings ohne geschlechterspezifische Trennung. Im fünften sind diese Merkmale noch als Dummy-Variable enthalten. Bei den Merkmalen der ersten Kategorie gibt es lediglich eine Einschränkung. Die Werte der drei Merkmalsträger mit den jeweils höchsten Ausprägungen wurden ersetzt durch die Durchschnittswerte ihrer jeweiligen Ausprägungen (punktuelle Mikroaggregation). So entsprechen die Maxima der Merkmale der ersten Kategorie nicht mehr den Originalwerten, sondern stellen die arithmetischen Mittel der drei höchsten Werte dar.

Neben der Informationsreduktion durch die Anonymisierung sind in die Datei zusätzlich generierte Informationen für die Nutzer aufgenommen worden. So wurde für Steuerpflichtige, die Einkünfte aus freien Berufen erzielen, aus der in der ursprünglichen Einkommensteuerstatistik vorliegenden Gewerbekennzahl das Merkmal „Freiberufler“ in den ersten vier Anonymisierungsbereichen generiert. Zusätzlich enthalten die Daten in allen Bereichen eine Dummy-Variable, die angibt, ob der Steuerpflichtige freiberuflich tätig ist.

Im Anonymisierungsbereich 5 sind die Merkmale der zweiten Kategorie nur noch als Dummy-Variable enthalten. Damit die Datennutzer die Struktur der Einkunftsentstehung auch im höchsten Einkommensbereich nachbilden können, wurden die sieben Einkunftsarten in drei Kategorien eingeteilt (Gewinneinkünfte, Einkünfte aus nichtselbständiger Arbeit und sonstige Überschusseinkünfte). Für jede dieser Kategorien existiert ein Bedeutungsmerkmal. Dieses nimmt den Wert 1 an, wenn in dieser Einkunftsart die höchsten Einkünfte erzielt werden und 3, wenn die geringsten Einkünfte aus dieser Kategorie stammen. Entsprechend weist dieses Merkmal die Ausprägung 2 für eine mittlere Bedeutung aus. Entstehen keine Einkünfte aus der Kategorie wird das Merkmal auf 0 gesetzt. Als Beispiel ist in Tabelle 2 die Häufigkeitsverteilung der Merkmale für die 1 000 Merkmalsträger mit den höchsten Einkommen angegeben.

**Tabelle 2: Struktur der Einkünfte bei den 1 000 Steuerpflichtigen mit den höchsten Einkommen**

Bedeutung	Gewinneinkünfte	Einkünfte aus nichtselbständiger Arbeit	Sonstige Überschusseinkünfte
Hoch	910	10	80
Mittel	49	318	616
Niedrig	30	275	283
Keine	11	397	21

Quelle: Merz, J., Vorgrimler, D., Zwick, M. (2004)

### **3 Faktisch anonymisierte Umsatzsteuerstatistik 2000**

#### **3.1 Datengrundlage<sup>7</sup>**

Im Rahmen der Umsatzbesteuerung werden von den Unternehmen Umsatzsteuer-Voranmeldungen bei den Finanzbehörden abgegeben. Durch die Auswertung der monatlichen und vierteljährlichen Umsatzsteuer-Voranmeldungen 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). Abweichungen zwischen den Angaben der Voranmeldungen und den tatsächlichen Umsätzen sind möglich, jedoch nicht von größerer Bedeutung.

In der Umsatzsteuerstatistik sind grundsätzlich alle Unternehmen abgebildet, die Lieferungen und Leistungen im Inland gegen Entgelt im Rahmen ihres Unternehmens erbringen. 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. Es ist zu beachten, dass bei finanziell, wirtschaftlich und organisatorisch zusammen gehörenden 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.

#### **3.2 Aussagekraft der Umsatzsteuerstatistik**

In der Umsatzsteuerstatistik werden u.a. folgende qualitative Merkmale abgebildet: Gewerkekennzahl, amtlicher Gemeindeschlüssel, Dauer der Steuerpflicht, Organschaft, Rechtsform. Die quantitativen Merkmale umfassen u. a.: steuerbarer Umsatz, Umsatzsteuer vor Abzug der Vorsteuerbeträge, abziehbare Vorsteuerbeträge, Umsatzsteuer-Vorauszahlung. Zum steuerbaren Umsatz der Unternehmen zählen im Umsatzsteuerrecht neben den als „Lieferungen und Leistungen“ bezeichneten Umsätzen der Unternehmen auch deren aus EU-Ländern importierte Güter („innergemeinschaftliche Erwerbe“). Als Umsatz werden im wirtschaftsstatistischen Sinn aber lediglich die Lieferungen und Leistungen betrachtet, da die innergemeinschaftlichen Erwerbe den Vorleistungen zuzurechnen sind.

Im Vergleich zu Primärerhebungen ist neben der erhebungstechnischen Abgrenzung des Umsatzes auch die inhaltliche Abgrenzung des steuerlichen Umsatzbegriffs zu beachten. Während in Primärerhebungen allein der Umsatz aus der laufenden Produktionstätigkeit betrachtet wird, beinhaltet der steuerliche Um-

---

<sup>7</sup> Ausführlich zur Umsatzsteuerstatistik vgl. Dittrich, S. (2004)

satz zusätzlich außerordentliche Erträge, z. B. aus einem nicht dem Betriebszweck dienenden Verkauf von Gebäuden.<sup>8</sup>

Das am häufigsten genutzte qualitative Merkmal ist die Wirtschaftszweigzuordnung (Klassifikation der Wirtschaftszweige, Ausgabe 1993). Maßgebend für die Zuordnung zu einem Wirtschaftszweig ist der Schwerpunkt der wirtschaftlichen Tätigkeit des Unternehmens.

### 3.3 Anonymisierungskonzeption

Um die faktische Anonymität der Merkmalsträger zu gewährleisten, mussten Informationen aus dem Originalmaterial der Umsatzsteuerstatistik vergrößert, gelöscht und in wenigen Fällen verändert werden. Der Einsatz von datenverändernden Verfahren wurde auf ein Minimum beschränkt. Im Folgenden wird die Anonymisierungskonzeption kurz vorgestellt.<sup>9</sup>

Die qualitativen Merkmale sind für den Datenschutz dann problematisch, wenn sie aufgrund einer tiefen Gliederung die Merkmalsträger in einer solchen Weise beschreiben, dass nur noch wenige dieser Beschreibung entsprechen (z. B. der einzige Bäcker einer Gemeinde). Um dies zu verhindern wurden folgende Maßnahmen durchgeführt:

- Die Regionalangabe wird auf Ost/West vergrößert, wobei der Osten die neuen Bundesländer inklusive Berlin umfasst.
- Die Wirtschaftszweigzuordnung wird in unterschiedlicher Tiefengliederung abhängig von den Besetzungszahlen mit z. T. neuen Positionen in die Daten aufgenommen.
- 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. € 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 Ausreißern datenschutzrechtliche Probleme. So ist z. B. besonders der Marktführer einer Branche gefährdet. Daher besteht das Anonymisierungskonzept bei den quantitativen Merkmalen aus einer zweistufigen Mikroaggregation. Die erste Stufe besteht aus einer für jedes Merkmal getrennt durchgeführten Mikroaggregation für alle Unternehmen. Dies gilt als die schwächste Form der Mikroaggregation.<sup>11</sup> Da durch diese getroffenen Maßnahmen

---

<sup>8</sup> Vgl. Treeck, H.-J. (2004)

<sup>9</sup> Ausführlich zur Anonymisierung vgl. Vorgrimler, D., Dittrich, S., Lenz, R., Rosemann, M. (2005a), zu Anonymisierungsverfahren allgemein vgl. Höhne, J. (2003)

<sup>10</sup> Bei über 150.000 Neugründungen trifft dies bei 118 Unternehmen zu.

<sup>11</sup> Zur Mikroaggregation vgl. Höhne, J. (2003), sowie Domnigo-Ferrer, J., Mateo-Sanz, J.M. (2002) und zur Wirkung der Mikroaggregation bei der Umsatzsteuerstatistik siehe Lenz, R., Vorgrimler, D. (2004)

die Marktführer noch nicht genügend geschützt werden, wird auf der zweiten Stufe eine punktuelle Mikroaggregation durchgeführt. Dabei werden nur speziell die jeweiligen drei regionalen Marktführer einer Branche mikroaggregiert.

Wie Analysen zur Schutzwirkung zeigten, ist durch die Durchführung der vorgestellten Maßnahmen die faktische Anonymität der Merkmalsträger gewährleistet.<sup>12</sup>

### 3.4 Beispielsanalysen aus der Umsatzsteuerstatistik

#### Exporttätigkeit der Unternehmen

Mit geringen Einschränkungen kann anhand des Merkmals der steuerfreien Lieferungen und Leistungen mit Vorsteuerabzug die Exporttätigkeit der Unternehmen beschrieben werden. Tabelle 3 zeigt die Exportquoten nach Umsatzgrößenklassen über alle Wirtschaftszweige, berechnet sowohl mit den Originaldaten als auch mit den anonymisierten Daten. Unterschiede in den Ergebnissen ergeben sich nur in geringem Maße. Es wird erwartungsgemäß eine deutlich stärkere Exportorientierung der Großunternehmen deutlich. Aber auch die mittelständischen Unternehmen der Größenklasse 25 bis unter 50 Mill. Euro erwirtschaften 16% ihrer Umsätze mit dem Ausland.

**Tabelle 3: Exportquoten nach Größenklassen**

Größenklasse der Lieferungen und Leistungen von ... bis unter ... EUR	Exportquote (%)	
	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

Quelle: Vorgrimler, D., Dittrich, S., Lenz, R., Rosemann, M. (2005b)

Neben diesen Auswertungen können die Exportquoten auch mit allen anderen qualitativen Merkmalen des Mikrodatenfiles kombiniert werden. Zudem lassen sich auch die EU-Exporte gezielt untersuchen.

#### Konzentrationsanalysen

Eine nahe liegende Untersuchungsmöglichkeit, welche die Umsatzsteuerstatistik bietet, ist die Analyse der Unternehmenskonzentration. Am einfachsten lässt sich die absolute Konzentration als Anteil der m größten Unternehmen am gesamten Merkmal berechnen (Konzentrationsrate  $CR_m$ , wobei m für die Anzahl der größten Unternehmen steht). Allerdings bleiben sämtliche Informationen außen vor, die über die anderen Unternehmen bekannt sind. Der Vorteil des Maßes liegt in der einfachen Berechenbarkeit. Ein Alter-

<sup>12</sup> Zur Analyse der Schutzwirkung vgl. Vorgrimler, D., Dittrich, S., Lenz, R., Rosemann, M. (2005a)

nativmaß ist der Herfindahlindex. Dieser Index ist gleich der Summe der quadrierten Marktanteile. Durch diese Definition werden Unternehmen mit hohen Marktanteilen überproportional gewichtet.

Da die Umsatzsteuerstatistik die Unternehmen fast vollständig erfasst, ist der Herfindahlindex problemlos zu berechnen. Schwierigkeiten bereitet allerdings die Anonymisierung der drei führenden Unternehmen einer Branche je Region, deren quantitativen Ausprägungen nur noch als Durchschnitt der drei führenden Unternehmen in den Daten enthalten sind. Während dies bei den Konzentrationsraten keinen Einfluss auf das Ergebnis hat, solange ein  $m$  von mindestens drei gewählt wird, führt sie beim Herfindahlindex zu einer systematischen Unterschätzung. Diese ist umso höher, je höher der Marktanteil des führenden Unternehmens im Verhältnis zu den nächst größeren ist.

**Tabelle 4: Beispiel für Konzentrationsmessung**

Branche	CR3		Herfindahl		Gini	
	original	anonymisiert	original	anonymisiert	original	anonymisiert
Datenverarbeitung und -banken	6,22	6,25	30,1	28,8	0,874	0,874
Forschung und Entwicklung	38,73	38,77	1.104,9	533,2	0,9142	0,9142

Quelle: Vorgrimler, D., Dittrich, S., Lenz, R., Rosemann, M. (2005b)

Beispielhaft wurden aus der Umsatzsteuerstatistik sowohl aus den original- als auch aus den anonymisierten Daten die beschriebenen absoluten Konzentrationsmaße für ausgewählte Branchen berechnet (siehe Tabelle 4). Während die Konzentrationsraten ( $m=3$ ) so gut wie keine Veränderungen aufgrund der Anonymisierung aufweisen, wird der Herfindahlindex wie erwartet bei hoher Konzentration unterschätzt. Hierbei dient der Bereich Forschung und Entwicklung als Extrembeispiel. Der Marktführer dieses Sektors ist neunmal größer als das nächstgrößte Unternehmen. Diese extreme Konzentration auf ein Unternehmen führt zu einer starken Unterschätzung des Herfindahlindex.

Der Gini-Koeffizient als weiteres Maß der Konzentration bleibt hingegen unverändert. Dies kann nicht weiter überraschen, da der Gini-Koeffizient sensibel gegenüber Veränderungen im mittleren Bereich der Verteilung reagiert, nicht aber gegenüber Veränderungen im oberen Bereich. Durch die Anonymisierung kam es allerdings lediglich zu Veränderungen im oberen Bereich.<sup>13</sup>

<sup>13</sup> Zur Kritik am Gini-Koeffizient vgl. Schaich, E. (1971). Weitere Beispielsanalysen zur Umsatzsteuerstatistik finden sich in Vorgrimler, D., Dittrich, S., Lenz, R., Rosemann, M. (2005b).

## 4 Entwicklungen beim Datenangebot

Standen in den bisherigen zwei Abschnitten bereits für die Wissenschaft bestehende Datenangebote im Mittelpunkt, werden im Folgenden neuere Entwicklungen in der amtlichen Steuerstatistik aufgezeigt. Diese sollen kurz- bzw. mittelfristig zu neuen Datenangeboten führen (zum Teil, wie im ersten Fall eines Scientific-Use-Files aus der Erbschaft- und Schenkungsteuerstatistik, steht die Veröffentlichung des Datenangebots unmittelbar bevor).

### 4.1 Faktisch anonymisierte Erbschaft- und Schenkungsteuerstatistik 2002

Erstmals seit 1978 wurde von der amtlichen Statistik für das Veranlagungsjahr 2002 wieder eine Erbschaft- und Schenkungsteuerstatistik veröffentlicht.<sup>14</sup> Sie wird zukünftig als dezentrale Erhebung fünfjährlich aufbereitet. In der Statistik erfasst sind alle Fälle, für die 2002 eine Steuer aufgrund einer Erbschaft oder Schenkung festgesetzt wurde. Wegen den relativ hohen Freibeträgen (für Ehegatten bei Erwerben von Todes wegen z.B. 307.000 €) unterliegt nur ein Teil dieser Vermögensübertragungen der Steuerpflicht und wird statistisch erfasst. Hierdurch wird das Volumen an Erbschaften und Schenkungen innerhalb eines Berichtszeitraums systematisch untererfasst.

Da diejenigen in der Statistik enthalten sind, die 2002 zur Steuer veranlagt wurden und zwischen der Vermögensübertragung und der Steuerfestsetzung zeitliche Verschiebungen entstehen können, ist die Statistik keine Abbildung eines Berichtsjahres. So kann es z.B. zu einem Erwerb von Todes wegen im Jahr 2000 kommen, die Steuer aber erst im Jahr 2002 festgesetzt werden. In einem solchen Fall ist der Steuerpflichtige in der Statistik enthalten. Umgekehrt ist nicht garantiert, dass falls jemand im Jahr 2002 erbt, er auch in der Statistik vertreten ist.

Aufgrund der beschriebenen Sachverhalte ist bei der Erstellung eines Scientific-Use-Files davon auszugehen, dass ein potenzieller Datenangreifer nur sehr ungenaue Informationen darüber hat, welche Merkmalsträger in der Statistik enthalten sind. Diese fehlende „Teilnahmekennntnis“ und zusätzlich die geringe Anzahl an Überschneidungsmerkmalen mit potenziellem Zusatzwissen<sup>15</sup> erlaubt es, eine Anonymisierungskonzeption zu wählen, die nur gering in das Datenmaterial eingreift. So wird bei dem Scientific-Use-File zum einen auf eine Stichprobe verzichtet. Zum anderen sind alle quantitativen Merkmale, gegliedert nach den Regionen Ost (neue Bundesländer inklusive Berlin) und West (alte Bundesländer ohne Berlin), unverändert enthalten. Einzige Ausnahme sind die jeweils drei höchsten Ausprägungen jedes Merkmals. Diese werden aus datenschutzrechtlichen Gründen analog zur Einkommensteuer- und zur Umsatzsteuerstatistik mit dem Mittelwert der drei höchsten Ausprägungen ersetzt (punktuelle Mikroaggregation).

---

<sup>14</sup> Zur neuen Erbschaft- und Schenkungsteuerstatistik allgemein und den wichtigsten Ergebnissen vgl. Zifonun, N., Schöffel, R. (2005).

<sup>15</sup> Das darüber hinaus nur schwer zu generieren ist.

#### 4.2 Public-Use-File aus der Lohn- und Einkommensteuerstatistik

Einzeldaten, deren Merkmalsträger in einer Form anonymisiert sind, dass eine Reidentifikation ausgeschlossen ist, dürfen vom Statistischen Bundesamt nach §16 Abs. 1 BStatG an Dritte übermittelt werden. Da in diesem Fall keine Einschränkung auf eine bestimmte Nutzergruppe vorliegt, werden Mikrodaten dieser Form „Public-Use-File“ genannt. Die hohen Anforderungen, die an die Geheimhaltung gestellt werden, bringen es mit sich, dass das Analysepotenzial von Public-Use-Files in der Regel sehr stark eingeschränkt ist. Daher sind diese Mikrodaten für wissenschaftliche Analysen in den meisten Fällen nicht von Interesse.

Dafür bietet ein Public-Use-File die Möglichkeit allen Nutzergruppen Einzeldaten – in diesem Fall Daten der Einkommensteuerstatistik – anzubieten. Als besonders interessante Zielgruppe erscheinen in diesem Zusammenhang Studenten der entsprechenden Fachrichtungen. Diese können Dank solcher Mikrodaten an Echtdaten ihre statistische Ausbildung voranbringen. Der geplante Public-Use-File aus der Lohn- und Einkommensteuerstatistik soll ein Angebot für diese Zielgruppe sein. Er soll Studenten die Möglichkeit bieten bereits in ihrer Ausbildung mit einer komplexen Erhebung statistisch zu arbeiten und darüber hinaus das Interesse an dem Thema „Einkommensteuerstatistik“ wecken. Aus diesem Grund wird der Public-Use-File in der Reihe „Campus-File“ der Forschungsdatenzentren des Bundes und der Länder erscheinen.<sup>16</sup>

Ausgangsmaterial für den Public-Use-File ist die aus dem Grundmaterial gezogene 1-% Stichprobe der Lohn- und Einkommensteuerstatistik 1998. Das Anonymisierungskonzept ist angelehnt an das Konzept zur faktischen Anonymisierung. Um allerdings die absolute Anonymisierung zu erreichen, werden innerhalb des vorgegebenen Rahmens deutlich stärkere Maßnahmen ergriffen. Dies beginnt bereits mit der erwähnten 1-% Stichprobe anstelle der bei FAST98 verwendeten 10-% Stichprobe. Die Anzahl der Merkmale wird darüber hinaus auf ungefähr 20 beschränkt und bei den Steuerpflichtigen mit einem GdE von über 1 Mill. € wird eine zusätzliche Unterstichprobe gezogen. Dies alles wird dazu führen, dass mit diesem Material keine anspruchsvolleren wissenschaftlichen Analysen möglich sein werden. Dies ist aber auch nicht das Ziel. Das Ziel Studenten die Möglichkeiten zu geben mit Echtdaten einfache aber interessante Auswertungen durchführen zu können wird mit diesem Material jedoch erreicht. So wird z. B. der Gesamtbetrag der Einkünfte repräsentativ in den Daten enthalten sein, womit Analysen zur Einkommensverteilung mit dem Public-Use-File möglich sein werden.

---

<sup>16</sup> Zu näheren Informationen zur Reihe „Campus-File“ vgl. Internet [www.forschungsdatennetzwerk.de/bestand/produkte/cf/index.asp](http://www.forschungsdatennetzwerk.de/bestand/produkte/cf/index.asp)

## 4.3 Jährliche Einkommensteuerstatistik

### 4.3.1 Abgrenzung zur amtlichen Bundesstatistik zur Lohn- und Einkommensteuer

Dem Statistischen Bundesamt wurde vom Bundesministerium der Finanzen die statistische Aufbereitung der Geschäftsstatistik zur Einkommensteuer ab dem Veranlagungsjahr 2001 übertragen (jährliche Einkommensteuerstatistik).<sup>17</sup> Dieses Datenmaterial umfasst jährlich ca. 28 Millionen Einkommensteuerveranlagungen und ab 2002 die Anträge zur Förderung der zusätzlichen Altersvorsorge (Riester Rente). Durch die Aufbereitung der Daten durch das Statistische Bundesamt wird es zukünftig möglich jährliche Auswertungen zur Struktur und Wirkungsweise der Einkommensteuer, zu deren wirtschaftlicher und sozialer Bedeutung und insbesondere zur Einkommens- und Steuerlastverteilung durchzuführen. Das umfangreiche Material wird sowohl die Politikberatung als auch die wissenschaftliche Forschung in vielen Bereichen erheblich verbessern. Die jährliche Einkommensteuerstatistik wird in der nächsten Zeit kontinuierlich auf den bereits in der Bundesstatistik bewährten Wegen allen Nutzergruppen zugänglich gemacht.

Die Länderfinanzverwaltungen übermitteln für die jährliche Einkommensteuerstatistik aus dem automatisierten Besteuerungsverfahren vorhandene Angaben jährlich auf Datenträgern. Diese Daten enthalten die im letzten Jahr in den Finanzämtern bearbeiteten Einkommensteuerveranlagungen, beschränkt auf die Einkommensteuerveranlagungen der drei vorangegangenen Jahre.<sup>18</sup> Im Statistischen Bundesamt werden diese Daten nach Veranlagungsjahren getrennt und aufbereitet. Nach Ende des dritten Kalenderjahres kann ein Veranlagungsjahr abgeschlossen werden. Das erste Veranlagungsjahr der Geschäftsstatistik ist das Jahr 2001, das Ende 2004 abgeschlossen werden konnte.

Durch den Wegfall der Zwischenjahre und die kontinuierliche Lieferung der Veranlagungsjahre verkürzt sich der time-lag, der zwischen der Beendigung eines Veranlagungsjahres und der Statistikerstellung liegt, auf zweierlei Wegen: Zum einen müssen bisher aufgrund der Dreijährlichkeit der amtlichen Bundesstatistik die Ergebnisse für drei Jahre die aktuellsten bleiben. Mit der jährlichen Geschäftsstatistik können dagegen die Ergebnisse jährlich aktualisiert und der time-lag um bis zu zwei Jahre reduziert werden. Zum anderen können für Veranlagungsjahre, deren regulärer Bearbeitungszeitraum von drei Jahren noch nicht beendet ist, bereits vorab erste Zwischenergebnisse erstellt werden. So können erste Tendenzen für das Veranlagungsjahr 2004 bereits nach der Datenlieferung zum Stichtag 31.12.2005 sichtbar werden. Für das Veranlagungsjahr 2003 sind bereits relativ stabile Ergebnisse zu erwarten. Gelingt es, auf Basis vorläufiger Daten das endgültige Veranlagungsergebnis zu prognostizieren, führt dies zu einer weiteren Reduzierung des time-lag.<sup>19</sup>

---

<sup>17</sup> Ausführlich zur jährlichen Einkommensteuerstatistik und zu den ersten Ergebnissen vgl. Lietmeyer, V., Kordsmeyer, V., Gräb, C., Vorgrimler, D. (2005), sowie Internet: [www.destatis.de/allg/d/veroe/proser4fist.htm](http://www.destatis.de/allg/d/veroe/proser4fist.htm)

<sup>18</sup> Zum Stand 31.12.2005 erhält das Statistische Bundesamt die in 2005 durchgeführten Veranlagungen für die Jahre 2002 bis 2004.

<sup>19</sup> Zur Problematik der Prognosen von Veranlagungsergebnissen vgl. Lietmeyer, V., Kordsmeyer, V., Gräb, C., Vorgrimler, D. (2005) S. 679.



Die jährliche Einkommensteuerstatistik weist neben der Periodizität noch weitere Unterschiede zur dreijährlichen amtlichen Bundesstatistik auf. Die wichtigsten können der Übersicht 3 entnommen werden.

**Übersicht 3: Die wichtigsten Unterschiede zwischen jährlicher Einkommensteuerstatistik und amtlicher Bundesstatistik über die Lohn- und Einkommensteuer**

	Statistik zur Einkommensteuer	
	Jährliche	Dreijährliche amtliche Bundesstatistik
<i>Grundgesamtheit:</i>		
Einkommensteuerveranlagungen	ja	ja
Nicht veranlagte Lohnsteuerkarten	nein	ja
Periodizität	jährlich	3-jährlich
Berichtsweg	zentral	dezentral
<i>Fiskalpolitische Aufgaben:</i>		
Gemeindeanteil an der Einkommensteuer	nein	ja
Lohnsteuererlegung	nein	ja
Plausibilitätsprüfungen	selektiv	ja
Regionale Gliederung	Land	Gemeinde

Quelle: Lietmeyer, V., Kordsmeyer, V., Gräß, C., Vorgrimler, D. (2005)

#### 4.3.2 Entwicklung eines Taxpayer-Panel

Neben der Möglichkeit den time-lag zwischen Beendigung eines Veranlagungszeitraums und Erstellung der Statistik zu verkürzen, bietet die Jährlichkeit der Statistik die Chance die Einkommensteuerstatistik zu einem Panel weiter zu entwickeln. Mit diesem können Steuerpflichtige über die Zeit beobachtet werden und so Anpassungsreaktionen auf Steuerrechtsänderungen und Änderungen sonstiger Rahmenbedingungen analysiert und daraus Schlüsse für die politische Beratung gezogen werden. Darüber hinaus lassen sich spezielle Untersuchungen wie Verteilungsanalysen auf Basis verschiedener Beobachtungszeiträume erstellen.

In der wissenschaftlichen Literatur hat sich mit der „new tax responsiveness literature“ eine eigene Forschungsrichtung etabliert, die steuerliche Fragestellungen auf Basis von Paneldaten analysiert. Ein erster deutscher Beitrag ist hierbei der von Gottfried/Schellhorn im Jahr 2003 erschienene Aufsatz „Die Elastizität des zu versteuernden Einkommens“. Diese Arbeit basiert auf einem Panel für Baden-Württemberg mit rund 50.000 Steuerfällen aus den Jahren 1988-1991. Für die deutsche Wissenschaft ist von erheblicher Bedeutung, dass für Forschungszwecke in diesem neuen empirischen Forschungszweig möglichst aktuelle und vollständige Daten zur Verfügung stehen.

Die prinzipiellen Möglichkeiten für die Verknüpfung gleicher Steuerpflichtiger über die Zeit sind dank der Jährlichkeit der Daten und der Steuernummer, als Identifikator der Merkmalsträger, gegeben. Problematisch ist ein eventueller Steuernummerwechsel eines Steuerpflichtigen. Die Finanzverwaltung sieht hierfür vor die alte Steuernummer separat mitzuliefern. Aus verschiedenen Gründen ist dies aber oftmals nicht möglich (so z. B. wenn ein Steuerpflichtiger über die Bundeslandgrenzen hinweg umzieht und das neu zuständige Finanzamt eine Steuernummer vergibt ohne die alte zu kennen). Daher wird das Problem des

Steuernummerwechsels nur teilweise gelöst. Die Zusammenführung der Veranlagungsjahre 2001 mit 2002 ergab 23,790 Mill. Steuerpflichtige. Hiervon konnten lediglich 45.000 Steuerpflichtige mit Hilfe der im Veranlagungsjahr 2002 als „alt“ geführten Steuernummer zusammengeführt werden. Tabelle 5 enthält die Anzahl an Steuerpflichtigen bei den verknüpften Daten und die Anzahl der jeweiligen verbliebenen Merkmalsträger. Dabei ist zu beachten, dass für Veranlagungsjahr 2002 und insbesondere für das Jahr 2003 mit der nächsten Datenlieferung weitere Daten geliefert werden, sodass erst danach ein korrekter Überblick über die nicht verknüpften Daten möglich ist.

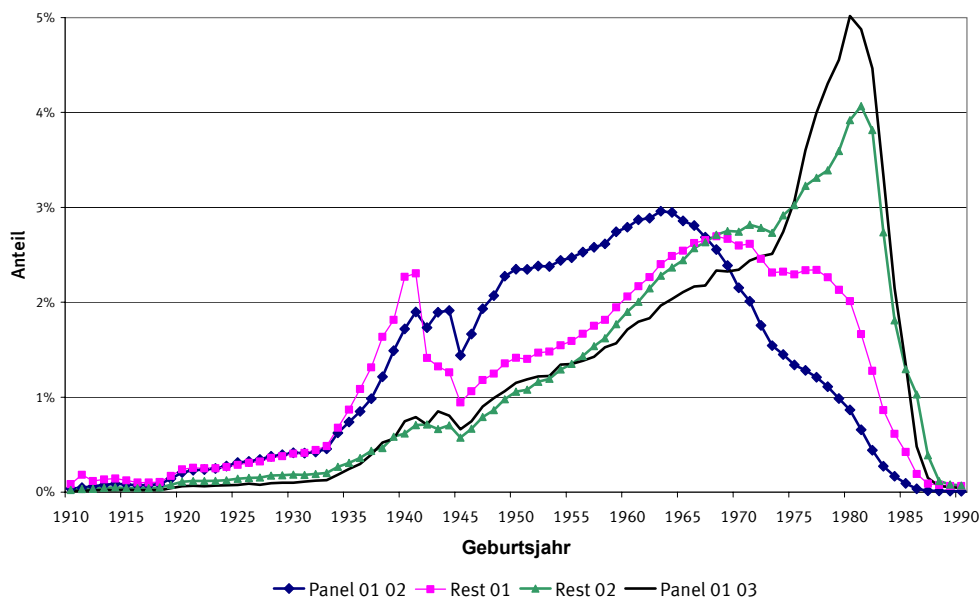
**Tabelle 5: Anzahl verknüpfter bzw. nicht verknüpfter Merkmalsträger**

Datei	Anzahl Merkmalsträger
Panel 01/02	23.790.324
Panel 01/02/03	13.444.950
Panel 01/03	182.912
Rest 01	3.919.057
Rest 02	2.713.979
Rest 03	3.191.671

Quelle: eigene Berechnungen

Panelmortalität kann entweder durch mangelnde Verknüpfbarkeit der Merkmalsträger über die Zeit aufgrund von fehlerhaften Identifikatoren oder aufgrund eines tatsächlichen Ausscheidens aus der bzw. Eintretens in die Population durch einen Merkmalsträger entstehen. Während es gilt, die erste Art der Mortalität zu verhindern, ist die zweite Art nicht nur nicht verhinderbar, sondern es kann für bestimmte Fragestellungen sogar erwünscht sein, diese Mortalität richtig zu identifizieren. Da beide Fälle auftreten können gilt es im ersten Schritt die beiden Mortalitätsarten bestmöglich voneinander abzugrenzen. Die Abbildung zeigt die Verteilung der Geburtsjahre bei verknüpften Merkmalsträgern (Panel 01/02) im Vergleich zu den Merkmalsträgern, die nicht verknüpft werden konnten. Merkmalsträger aus dem Veranlagungsjahr 2001 können u. a. dann nicht verknüpft werden, wenn sie 2002 keine Einkommensteuererklärung mehr abgegeben haben. Gründe hierfür wären das Erreichen des Rentenalters oder der Tod. Dagegen können die restlichen Merkmalsträger aus dem Veranlagungsjahr 2002 erstmalig eine Einkommensteuererklärung abgegeben haben und daher nicht mit 2001 verknüpfbar sein. Während der erste Fall hauptsächlich auf ältere Steuerpflichtige zutrifft, trifft der zweite tendenziell bei jüngeren Steuerpflichtigen zu. Daher ist bei der Verteilung der Geburtsjahre für die restlichen Merkmalsträger des Jahres 2001 eine Verschiebung zu den älteren Geburtsjahrgängen zu erwarten und für die des Jahres 2002 zu den jüngeren. Die Abbildung bestätigt diese Vermutung. Für die restlichen Merkmalsträger aus 2001 ergibt sich eine starke Ausprägung bei den Jahrgängen, die typischerweise in Rente gegangen sind. Darüber hinaus sind die sehr alten Jahrgänge überproportional vertreten. Dies kommt aufgrund des Maßstabes der Abbildung nicht zur Geltung. Bei den restlichen Merkmalsträgern aus 2002 sind dagegen die jüngeren Jahrgänge überproportional vertreten. Dieses Ergebnis zeigt, dass ein noch zu bestimmender Anteil an den Ausfällen auf tatsächliche Ausfälle zurück zu führen ist.

**Abbildung: Verteilung der Geburtsjahre bei den verknüpften Merkmalsträgern und den nicht verknüpften Merkmalsträgern im Vergleich**



Auf einen Sonderaspekt weisen die verknüpften Merkmalsträger aus dem Panel 01/03 hin. Diese enthalten Merkmalsträger des Veranlagungsjahres 2001 die nicht mit Merkmalsträgern des Veranlagungsjahres 2002 jedoch mit Merkmalsträgern aus dem Jahr 2003 verknüpft werden konnten. Die Steuerpflichtigen haben demnach ein Jahr mit der Veranlagung ausgesetzt. Die Struktur der Geburtsjahre zeigt eine starke Konzentration bei den jungen Jahrgängen. Diese Steuerpflichtigen weisen relativ unstetige Erwerbsverläufe auf, was tendenziell eher bei jüngeren Steuerpflichtigen in der Ausbildung bzw. Studium der Fall sein dürfte.

Die Entwicklung eines Taxpayer-Panels steht noch am Anfang. Fragen wie die der Datenhaltung, von Verknüpfungsstrategien und dem Umgang mit der Panelmortalität bedürfen noch einer näheren Untersuchung. Erste Ergebnisse zeigen, dass die Möglichkeit eines Panels für die jährliche Einkommensteuerstatistik prinzipiell besteht. Nicht zuletzt muss jedoch noch eine ausreichende Datenbasis aufgebaut werden um überhaupt ein aussagefähiges Panel aufbauen zu können. Mit der nächsten Datenlieferung zur jährlichen Einkommensteuerstatistik im Frühjahr 2006 wird hierfür ein wichtiger Schritt getan werden können.

#### 4.4 Panel aus der Umsatzsteuerstatistik

Im Rahmen des Projektes „Wirtschaftsstatistische Paneldaten und faktische Anonymisierung“ werden Panels von verschiedenen amtlichen wirtschaftsstatistischen Einzeldaten entwickelt. Die Umsatzsteuerstatistik ist hierbei eine der Projektstatistiken. Sie soll für die Jahre 2000 bis 2004 zu einem Panel verknüpft werden. Dieses soll der Wissenschaft nicht erst nach Ende der dreijährigen Projektlaufzeit zur Verfügung stehen, sondern über Fernrechnen und/oder Sonderaufbereitung bereits unmittelbar nach der erfolgreichen

Entwicklung, für die etwa ein Jahr angesetzt ist. Am Ende des Projektes soll neben dem innerhalb der amtlichen Statistik zugänglichem Panel ein faktisch anonymisiertes Panel vorliegen, das der Wissenschaft als Scientific-Use-File außerhalb der amtlichen Statistik zur Verfügung gestellt werden kann.

Ein Panel der Umsatzsteuerstatistik eröffnet der amtlichen Statistik und der externen Wissenschaft neue Analysemöglichkeiten z.B. im Bereich der dynamischen Konzentration.<sup>20</sup> Durch den bisher angebotenen Scientific-Use-File 2000 sind lediglich statische Betrachtungen der Konzentration möglich. Die Umsatzsteuerstatistik kann sich unter bestimmten Umständen auch für Analysen der Unternehmensdemographie eignen. Im Scientific-Use-File 2000 ist das Merkmal Unternehmensneugründung enthalten. Aufgrund mangelnder Qualität ist es bisher nicht vertretbar gewesen, auch die Unternehmenslöschungen mit in die Daten aufzunehmen.<sup>21</sup> Gelingt es, die Qualität der Ausprägung „Neugründung“ und „Löschung“ evtl. mit Hilfe weiterer statistischer Quellen ausreichend zu verbessern, können im Bereich der Unternehmensdemographie wertvolle Analysemöglichkeiten entstehen.

## 5 Fazit

Mit den Scientific-Use-Files zur Lohn- und Einkommensteuerstatistik 1998 und zur Umsatzsteuerstatistik 2000 wurde die steuerstatistische Datenbasis für die Wissenschaft bereits erheblich verbessert. Darüber hinaus wird – wie der Beitrag gezeigt hat – die amtliche Steuerstatistik ihr Datenangebot für die Wissenschaft kontinuierlich ausbauen. Die weiteren geplanten Scientific- und Public-Use-Files werden für weitere Forschungsimpulse sorgen. Der Aufbau von Paneldaten als zusätzliches Datenangebot wird neue Analysemöglichkeiten eröffnen und bisherige Analysen auf eine breitere Basis stellen.

Die Einschränkungen des Analysepotenzials aufgrund der getroffenen Anonymisierungsmaßnahmen bei den einzelnen Scientific-Use-Files bringt es mit sich, dass nicht alle wissenschaftlichen Analysen in gewünschter Form möglich sind bzw. sein werden. In diesen Fällen stehen den Wissenschaftlern über die Forschungsdatenzentren des Bundes und der Länder in Zusammenarbeit mit der Fachabteilung alternative Zugangswege zu den Mikrodaten zur Verfügung. Die Fachabteilung begreift sich sowohl bei der Nutzung der Scientific-Use-Files als auch bei dem Zugang über die alternativen Zugangswege als Partner der Wissenschaft zur Unterstützung von Forschungsarbeiten.

---

<sup>20</sup> Zur dynamischen Konzentrationsmessung vgl. Vorgrimler, D. (2002), S. 332 ff..

<sup>21</sup> Zu den Möglichkeiten mit der Umsatzsteuerstatistik auch Unternehmenslöschungen zu betrachten vgl. Treeck, H.-J. (2004), S. 8 ff.

## Literaturhinweise

*Bach, S., Haan, P., Rudolph, H.-J., Steiner, V.* (2004): Reformkonzepte zur Einkommens- und Ertragsbesteuerung: Erhebliche Aufkommens- und Verteilungswirkungen, aber relativ geringe Effekte auf das Arbeitsangebot, in: DIW-Wochenbericht 16.

*Dittrich, S.* (2004): Umsatz und ihre Besteuerung 2002, in: Wirtschaft und Statistik, Heft 6, Seite 1195-1200.

*Domnigo-Ferrer, J., Mateo-Sanz, J.M.* (2002): Practical data-oriented microaggregation for statistical disclosure control, IEE Transaction on Knowledge and Data Engineering, Vol. 14(1), Seite 189ff.

*Gottfried, P., Schellhorn, H.* (2003): Die Elastizität des zu versteuernden Einkommens, IAW-Diskussionspapier Nr. 14.

*Höhne, J.* (2003): Methoden zur Anonymisierung wirtschaftsstatistischer Einzeldaten, in: Gnos/Ronning, G. (Hrsg.): Anonymisierung wirtschaftsstatistischer Einzeldaten“, Band 42 der Schriftenreihe „Forum der Bundesstatistik“, Wiesbaden, Seite 77-80.

*Höhne, J., Sturm, R., Vorgrimler D.* (2003): Konzept zur Schutzwirkung faktischer Anonymisierung, in: Wirtschaft und Statistik, Heft 4, S. 287-292.

*Lenz, R., Vorgrimler, D.* (2004): Geheimhaltungsmethoden auf dem Prüfstand – eine Analyse anhand der Umsatzsteuerstatistik, in: Wirtschaft und Statistik, Heft 6, Seite 639- 648.

*Lenz, R., Sturm, R., Vorgrimler, D.* (2004): Maße für die faktische Anonymität von Mikrodaten, in: Wirtschaft und Statistik, Heft 6, Seite 621-638.

*Lietmeyer, V., Kordsmeyer, V., Gräß, C., Vorgrimler, D.* (2005): Jährliche Einkommensteuerstatistik auf Basis der bisherigen Geschäftsstatistik der Finanzverwaltung, in: Wirtschaft und Statistik, Heft 7, Seite 671-681.

*Merz, J., Vorgrimler, D., Zwick, M.* (2004): Faktisch anonymisiertes Mikrodatenfile der Lohn- und Einkommensteuerstatistik 1998, in: Wirtschaft und Statistik, Heft 10, Seite 1079-1091.

*Rosinus, W.* (2000): Die steuerliche Einkommensverteilung, in Wirtschaft und Statistik, Heft 6, Seite 456-463.

*Schaich, E.* (1971): Lorenzfunktion und Gini-Koeffizient in kritischer Betrachtung, in: Jahrbücher für Nationalökonomie und Statistik, Band 185, Seite 193-207.

*Treeck, H.J.* (2004): Die Umsatzsteuerstatistik als Quelle wirtschaftsstatistischer Analysen, in: Statistische Analysen und Studien NRW, Band 15, Seite 3-10.

*Vorgrimler, D.* (2002): Anbieterkonzentration auf dem Agrartechnikmarkt am Beispiel des Traktoren- und Mähdreschermarktes, in: *Agrarwirtschaft*, Heft 7, Seite 332-340.

*Vorgrimler, D., Dittrich, S., Lenz, R., Rosemann, R.* (2005a): Ein Scientific-Use-File der Umsatzsteuerstatistik 2000, in: *Wirtschaft und Statistik*, Heft 3, Seite 201-210.

*Vorgrimler, D., Dittrich, S., Lenz, R., Rosemann, R.* (2005b): Wissenschaftliche Analysen mit Hilfe der amtlichen Umsatzsteuerstatistik, in: *Wirtschaftswissenschaftliches Studium*, Heft 10, Seite 527-531.

*Zifonun, N., Schöffel, R.* (2004): Neue Erbschaft- und Schenkungsteuerstatistik 2002, in: *Wirtschaft und Statistik*, Heft 9, Seite 1028-1035.