

**Absolute Anonymisierung in der Steuerstatistik
–CAMPUS-File der Einkommensteuer 1998 –**

1 Einführung

Die Möglichkeit der Weitergabe von Einzeldaten aus der amtlichen Statistik ist in § 16 des Gesetzes über die Statistik für Bundeszwecke (Bundesstatistikgesetz, BStatG) geregelt. Nach § 16 Abs. 1 Satz 2 Nr. 4 dürfen Einzeldaten beispielsweise dann weitergegeben werden, „wenn sie dem Befragten oder Betroffenen nicht zuzuordnen sind.“

Mit der Veröffentlichung des Campus-File bietet das Statistische Bundesamt erstmals frei zugängliche Einzeldaten aus der Einkommensteuerstatistik, anhand derer das Besteuerungsverfahren nachvollzogen und Simulationen durchgeführt werden können.

2 Anonymisierung

Die anonymisierte Datei basiert auf einer 1%-Stichprobe der Einkommensteuerstatistik, wobei sich die eingesetzten Anonymisierungsmaßnahmen auf die traditionellen Anonymisierungsmethoden beschränken.¹

Anhand ihres Gesamtbetrags der Einkünfte (GdE) werden die Steuerpflichtigen zu Anonymisierungszwecken in die in Tabelle 1 dargestellten Bereiche eingeteilt:

Tabelle 1: Einteilung der Anonymisierungsbereiche

Anonymisierungsbereich	GdE in € (DM)
1	bis 511.291 (bis 1.000.000)
2	über 511.291 (über 1.000.000)
3	top 10
4	bottom 10

Merkmalskategorien

Die stetigen Merkmale sind nach ihrer Bedeutung in drei Kategorien eingeteilt. In der ersten sind die Merkmale enthalten, die unverändert übernommen werden. Die Merkmale der zweiten Kategorie sind zu Dummy-Variablen transformiert, die den Wert 1 annehmen, wenn das Merkmal vorhanden ist. Alle weiteren Merkmale gehören zur dritten Kategorie und werden gelöscht.

¹ Zu den Methoden vgl. Höhne, J., Methoden zur Anonymisierung wirtschaftsstatistischer Einzeldaten, in: Ronning, G., Gnos, R., Anonymisierung wirtschaftsstatistischer Einzeldaten, 2003, Forum der Bundesstatistik Band 42, S. 69-94.

Merkmale der ersten Kategorie:

- Summe der Einkünfte (A+B)²
- Gesamtbetrag der Einkünfte
- Einkommen
- zu versteuerndes Einkommen
- tarifliche Einkommensteuer
- festzusetzende Einkommensteuer

Merkmale der zweiten Kategorie:

- die sieben Einkunftsarten, getrennt nach A und B
- Sonderausgaben, die nicht Vorsorgeaufwendungen sind
- Sonderausgaben: Vorsorgeaufwendungen
- Außergewöhnliche Belastungen, abzugfähig – bei getrennter Veranlagung –A –
- Außergewöhnliche Belastungen, abzugfähig – bei getrennter Veranlagung –B –
- Förderung des Wohneigentums: Steuerbegünstigungen insgesamt

Die qualitativen Merkmale wurden stark vergrößert oder teilweise weggelassen. Tabelle 2 fasst die in der Datei vorhandenen Merkmale und ihre Ausprägungen zusammen:

Tabelle 2: Merkmale und ihre Ausprägungen

Merkmal		Ausprägung
Merker		1 = veranlagte Fälle 2 = manuelle Fälle
Veranlagungsart		1 = Grundtabelle 2 = Splittingtabelle
Kinder		Ja/nein
Alter		Klassen mit 10 Jahren (untere Klasse < 20, obere Klasse >= 70)
Region		West/Ost
stetige Merkmale	Kategorie 1	Ja
	Kategorie 2	Dummy
	Kategorie 3	Nein

Zusätzlich werden in den Anonymisierungsbereichen 3 und 4 die Ausprägungen der zehn Merkmalsträger mit dem höchsten und dem geringsten GdE durch die Durchschnittswerte ihrer jeweiligen Ausprägungen ersetzt. So entsprechen die Maxima und Minima der Merkmale der ersten Kategorie nicht mehr den Originalwerten, sondern stellen die arithmetischen Mittel der zehn höchsten bzw. geringsten Werte dar.

3 Fazit

Mit der vorgelegten Anonymisierungskonzeption ist es gelungen, erstmalig absolut anonymisierte Daten aus der Einkommensteuerstatistik zu erstellen, die einem breiten Nutzerkreis zu Analysezwecken zur Verfügung gestellt werden können. Da das Analysepotenzial durch die Anonymi-

² A bezeichnet männliche Steuerpflichtige, B weibliche.

sierung eingeschränkt ist, sind nicht alle Fragestellungen mit den Daten analysierbar. Für die Wissenschaft sei deshalb auf die alternativen Zugangswege zu Mikrodaten verwiesen, die von den Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder angeboten werden (<http://www.forschungsdatenzentrum.de>).

Trotz dieser Einschränkung stellt die angebotene Datei einen großen Fortschritt für die steuerstatistische Datenbasis dar. Erstmals ist ein so umfangreiches Material für steuerstatistische Analysen allgemein und kostenlos zugänglich. Insbesondere stellt das CAMPUS-File Informationen über die Bezieher der höchsten Einkommen bereit. Gerade dies ist ein Bereich, der in anderen Datenbasen entweder nicht oder nur sehr unzureichend abgebildet ist.

Wiesbaden, im November 2005

Literatur

Höhne, J.: Methoden zur Anonymisierung wirtschaftsstatistischer Einzeldaten in: Ronning, G., Gnoss, R., Anonymisierung wirtschaftsstatistischer Einzeldaten, 2003, Forum der Bundesstatistik, Band 42, S. 69-94.

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

Köhler, S.: Anonymisierung von Mikrodaten in der Bundesstatistik und ihre Nutzung – Ein Überblick, in: *Forum der Bundesstatistik*, Band 31, 1999, S. 133-150.

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

Zwick, M.: Einzeldatenmaterial und Stichproben innerhalb der Steuerstatistik, in: *Wirtschaft und Statistik*, Heft 7, 1998, S. 566-572.

Ansprechpartner:

Dr. Nicole Buschle

Christopher Gräb

E-Mail: nicole.buschle@destatis.de

E-Mail: christopher.graeb@destatis.de

Weitere Informationen

http://www.destatis.de/themen/d/thm_finanzen.php

<http://www.forschungsdatenzentrum.de>