FDZ-Arbeitspapier
Nr.1

Heiko Müller

STATISTISCHE ÄMTER
DES BUNDES UND DER LÄNDER
FORSCHUNGSDATENZENTREN

Ein Vergleich der Ergebnisse von Mikrosimulationen mit denen von Gruppensimulationen auf Basis der Einkommensteuerstatistik

2005

FDZ-Arbeitspapier
Nr.1

Heiko Müller

STATISTISCHE ÄMTER
DES BUNDES UND DER LÄNDER
FORSCHUNGSDATENZENTREN

Ein Vergleich der Ergebnisse von Mikrosimulationen mit denen von Gruppensimulationen auf Basis der Einkommensteuerstatistik

2005

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 März 2005

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 angeboten.

© Statistisches Bundesamt, Wiesbaden 2005
(im Auftrag der Herausgebergemeinschaft)

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

Ein Vergleich der Ergebnisse von Mikrosimulationen mit denen von Gruppensimulationen auf Basis der Einkommensteuerstatistik

Heiko Müller[*]

I. Einleitung

Mit der Novellierung des Gesetzes über Steuerstatistiken (StStaG) im Jahr 1996 wurde es erstmalig möglich, die Einzeldatensätze der Einkommensteuerstatistik zentral zusammenzuführen und für Zusatz- und Sonderaufbereitungen zu nutzen.[1] Dadurch sind die Statistischen Ämter des Bundes und der Länder in der Lage, Daten flexibel aufzubereiten und Auswertungen zu erstellen, die über die mit dem speziell für die Lohnsteuer- und Einkommensteuerstatistik entwickelten Tabellenerstellungsprogramm (TEP) erzeugten Standardtabellen hinausgehen. Zudem können Fragestellungen, deren Beantwortung anhand der veröffentlichten aggregierten Daten der Einkommensteuerstatistik nur schwer möglich oder gar unmöglich ist, mittels vom Statistischen Bundesamt durchgeführter Mikrosimulationsrechnungen untersucht werden.

Jedoch kann bislang insbesondere bei Ad-hoc-Untersuchungen zur Änderung des Aufkommens der Einkommensteuer sowohl auf Grund des beschränkten Zugangs zu den relevanten Daten als auch aus Kostengründen nicht immer auf die Einzeldaten der Einkommensteuerstatistik zurückgegriffen werden.[2] Dagegen liegen die Ergebnisse der Einkommensteuerstatistik zusammengefasst in Tabellen als klassierte Daten vor, die u. a. in der vom Statistischen Bundesamt herausgegebenen Fachserie 14 „Finanzen und Steuern“ veröffentlicht werden. Im vorliegenden Beitrag werden die Ergebnisse von Mikrosimulationen auf Basis von Einzeldatensätzen der Einkommensteuerstatistik mit den Ergebnissen von Gruppensimulationen auf Grundlage klassierter Daten der Einkommensteuerstatistik verglichen. Anhand der Ergebnisse werden die Einsatzmöglichkeiten und die Grenzen des verwendeten Gruppensimulationsmodell aufgezeigt.

Die dem Beitrag zu Grunde liegenden Berechnungen wurden im Rahmen der Dissertation des Autors mit dem Titel „Das Aufkommen der Steuern vom Einkommen in Deutschland Anfang und Mitte der neunziger Jahre – Eine empirische Untersuchung der Gründe für die vom Volkseinkommen abweichende Entwicklung des Aufkommens der Steuern vom Einkommen“ durchgeführt. Diese Arbeit entstand im Zusammenhang mit einem von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekt zur Erforschung der Ursachen für das Sinken des Aufkommens der veranlagten Einkommensteuer und der Körperschaftsteuer.

* Dipl. Kaufmann Dr. Heiko Müller ist wissenschaftlicher Mitarbeiter des Instituts für Betriebswirtschaftliche Steuerlehre der Humboldt Universität Berlin. Bei den dargestellten statistischen Angaben handelt es sich um eigene Arbeitsergebnisse des Autors im Zusammenhang mit der Nutzung der bereitgestellten Daten der Forschungsdatenzentren. Der Beitrag liegt in der alleinigen Verantwortung des Autors.

[1] Vgl. *Zwick*, WiSta 1998, S. 566 (570).

[2] Mit der Einrichtung der Forschungsdatenzentren (siehe auch www.forschungsdatenzentrum.de) hat sich der Zugriff auf die Daten der amtlichen Statistik verbessert.

Die am Institut für Betriebswirtschaftliche Steuerlehre der Humboldt-Universität zu Berlin durchgeführten empirischen Untersuchungen zum Aufkommen der Einkommen- und Körperschaftsteuer erfolgten in enger Kooperation mit dem Statistischen Bundesamt. Es war das erste Projekt bei dem das Statistische Bundesamt Sonderauswertungen für eine Universität auf Basis der Einzeldatensätze der amtlichen Steuerstatistik erstellte.[3] Für die Untersuchungen haben die vom Statistischen Bundesamt durchgeführten Mikrosimulationsrechnungen und erstellten Sonderauswertungen einen wesentlichen Beitrag geleistet.

II. Die Steuerstatistik des Statistischen Bundesamtes als Datenbasis

Die Güte empirischer Untersuchungen hängt in einem ganz entscheidenden Maße vom Umfang und der Qualität der verfügbaren Daten ab. Für Simulationsrechnungen zum Aufkommen der Einkommensteuer müssen Daten zur Entstehung und Verteilung der Einkünfte, der steuerlichen Bemessungsgrundlagen und der Einkommensteuer herangezogen werden. Dafür stehen in der Bundesrepublik Deutschland grundsätzlich mehrere potenzielle Datenquellen zur Verfügung.[4]

Für eine insbesondere auf steuerrechtsinduzierte Aufkommenswirkungen fokussierte Simulation der Besteuerung des Einkommens stellt die amtliche Einkommensteuerstatistik sowohl hinsichtlich der Vollständigkeit und Detailliertheit der steuerrelevanten Daten als auch bezüglich der Verfügbarkeit die beste Datenbasis dar.

Bei der Einkommensteuerstatistik [5] handelt es sich um eine Sekundärstatistik, d.h., die Finanzbehörden stellen für die Steuerstatistik Daten zur Verfügung, die im Rahmen des Steuerveranlagungs- und Steuerfestsetzungsverfahrens anfallen. Es handelt sich insoweit um authentische Daten. Die Angaben werden also nicht mittels Befragung erhoben, sondern es sind die Sachverhalte, die auf Grundlage der Steuererklärungen von der Finanzverwaltung festgestellt und bei der Veranlagung tatsächlich einbezogen wurden. Die Einkommensteuerstatistik wird seit 1965 jedoch nur in einem Dreijahresturnus erstellt[6] und erscheint erst mit einem zeitlichen Abstand zum relevanten Veranlagungszeitraum von mindestens vier Jahren. Auf Grund des Erhebungsturnus sind lückenlose intertemporale Analysen auf Basis der Steuerstatistik leider nicht möglich.[7] Die letzte, 1999 veröffentlichte Einkommensteuerstatistik liegt für den Veranlagungszeitraum 1995 vor.

---

[3] Für die sehr gute Zusammenarbeit möchte ich mich insbesondere bei dem jetzigen Leiter des FDZ des Statistischen Bundesamts, Herrn Diplom-Volkswirt Markus Zwick, sowie seiner für den Bereich Datenmanagement, Sonderaufbereitungen und Ferndatenzugriff verantwortlichen Mitarbeiterin, Frau Diplom-Betriebswirtin Heike Habla, recht herzlich bedanken.

[4] Hier sind insbesondere die vom Statistischen Bundesamt erstellte Volkswirtschaftliche Gesamtrechnung, die amtliche Steuerstatistik und die Einkommens- und Verbrauchsstichprobe (EVS) zu nennen (vgl. u.a. *Euler*, WiSta 1985, S. 56). Weiterhin existieren einige im Rahmen von Forschungsprojekten entwickelte Datenbasen, wie z.B. die Einkommens- und Transferschichtung und das sozio-ökonomische Panel des Deutschen Instituts für Wirtschaftsforschung (DIW) oder die Mikrodaten des Instituts für angewandte Wirtschaftsforschung (IAW). Zur Problematik geeigneter Datensätze für steuerpolitische Simulationen siehe ausführlich *Spahn* et al., Mikrosimulation in der Steuerpolitik, 1992, S. 116 – 125, *Bork*/*Petersen*, Ein Vergleich möglicher Datensätze zur Eignung für steuerpolitische Simulationsrechnungen; Finanzwissenschaftliche Diskussionsbeiträge Nr. 20, Universität Potsdam 1997, und *Bork*, Steuern, Transfers und private Haushalte – Eine mikroanalytische Simulationsstudie der Aufkommens- und Verteilungswirkungen, 2000, S. 91 – 112.

[5] Bis zum Veranlagungszeitraum 1989 wurde separat von der Einkommensteuerstatistik, die bis dahin nur die veranlagten Einkommensteuerpflichtigen enthielt, eine Lohnsteuerstatistik ausgewiesen. Da die Zusammenführung der zwei Steuerstatistiken problematisch war, stellte die am Erhebungsverfahren orientierte Trennung insbesondere für Aufkommens- und Verteilungsanalysen ein Hemmnis dar. So auch *Krupp*, Möglichkeiten der Verbesserung der Einkommens- und Vermögensstatistik, 1975, S. 61. Mit der Novellierung des StStaG im Rahmen des Jahressteuergesetzes 1996 vom 11.10.1995 (BGBl. I S. 1250) erfolgte für den Veranlagungszeitraum 1992 eine Zusammenführung der Statistiken (siehe dazu auch *von der Lippe*, SteuerStud 1997, S. 265). Das Statistische Bundesamt verzichtete für den Veranlagungszeitraum 1995 gänzlich auf die Veröffentlichung einer eigenständigen Lohnsteuerstatistik im Rahmen der Fachserie 14 „Finanzen und Steuern".

[6] Vor der Einführung des dreijährigen Erhebungsturnus mit dem Inkrafttreten des StStaG vom 6.12.1966 (BGBl. I S. 665) wurde die Steuerstatistik seit ihrer Wiederaufnahme nach dem Zweiten Weltkrieg im Jahre 1950 nur in unregelmäßigen Zeitabständen durchgeführt. Siehe dazu auch *Fuß*, Die Steuerstatistiken über das Einkommen; in: Stand der Einkommensteuerstatistik, Individual- und Haushaltseinkommen, Einkommensschichtung, Hrsg. *Fürst*, 1974, S. 23 (24).

[7] Das Statistische Bundesamt ist jedoch bemüht, das Problem des Mangels an aktuellen und vollständigen Daten, die insbesondere von der Wissenschaft und Politik benötigt werden, zu lösen. Durch das Altersvermögensgesetz vom 26.6.2001 (BGBl. I S. 1310) erfolgte eine Änderung des StStaG dahin gehend, dass ab dem Veranla-

Die lange Aufbereitungszeit der Steuerstatistik hat verschiedene Ursachen. Die Finanzverwaltungen der Bundesländer haben den statistischen Landesämtern innerhalb einer Frist von zwei Jahren und neun Monaten die entsprechenden Informationen aus dem Veranlagungsverfahren in Form von Datensätzen auf maschinellen Datenträgern bereitzustellen.[8] Dieser relativ lange Zeitraum ist neben der meist starken Arbeitsbelastung der Finanzämter auch durch die gesetzlichen Steuererklärungsfristen[9] und die Festsetzungsfristen[10] bedingt, um die Vollständigkeit und Korrektheit der Steuerstatistik zu gewährleisten.[11] Die statistischen Landesämter unterwerfen die von den Finanzverwaltungen übermittelten Daten einer statistisch differenzierten Plausibilitätsprüfung und bereinigen diese gegebenenfalls.[12] Spätestens nach einem halben Jahr sollen dann die aufbereiteten Daten der Länder an das Statistische Bundesamt übermittelt werden. Das Statistische Bundesamt führt die Ergebnisse der einzelnen Bundesländer zusammen, wobei zeitliche Verzögerungen auftreten können, da die Verwaltungen und Informationsdienste nicht einheitlich aufgebaut sind.[13] Die Erstellung einer auf früheren Statistiken basierenden Hochrechnung zu einem früheren Zeitpunkt wird insbesondere wegen der zumeist erst spät bearbeiteten „komplizierten, gewichtigen“[14] Veranlagungen als problematisch erachtet.[15] Jedoch verkürzte sich in den letzten Jahren die Aufbereitungszeit für die Steuerstatistik, da mit der fortschreitenden Entwicklung der elektronischen Datenverarbeitung die Verwaltung und Bearbeitung der sehr umfangreichen Daten der Einkommensteuerstatistik erheblich vereinfacht und flexibilisiert werden konnten.[16]

Die in der Einkommensteuerstatistik ausgewiesene festgesetzte Steuer entspricht nicht dem kassenmäßigen Aufkommen des entsprechenden Veranlagungszeitraumes. Dies ist zum einen darin begründet, dass in der Einkommensteuerstatistik die Lohnsteuer der nicht an das Finanzamt zurückgegebenen Lohnsteuerkarten, die pauschale Lohnsteuer sowie die übrigen nichtveranlagten Steuern vom Einkommen – insbesondere die Kapitalertragsteuer der nichtveranlagten Steuerpflichtigen – nicht erfasst werden. Zum anderen fallen die Entstehung der Steuerschuld und ihre Begleichung zeitlich auseinander. Dieser Timelag ist für jede Steuerart und Steuererhebungsform unterschiedlich und wandelt sich im Zeitablauf, da sich eine Vielzahl von Einflussfaktoren ändert.[17]

Die Erklärungs-, Veranlagungs- und Zahlungsverfahren werden sowohl von rechtlichen und institutionellen Faktoren als auch vom Verhalten der Steuerpflichtigen beeinflusst. So ergeben sich z.B. durch Rechtsänderungen Umstellungen bei den Steuererhebungsformen und -verfahren, aber auch die von den Steuer-

---

gungszeitraum 2001 das Statistische Bundesamt die jährliche Geschäftsstatistik zur Einkommensteuer des BMF aufbereitet und somit ab 2005 auch diese Einzeldaten für wissenschaftliche Untersuchungen prinzipiell zur Verfügung stehen. Zu den Möglichkeiten der wissenschaftlichen Nutzung der Steuerstatistik vgl. *Zwick*, ZWS 2001, S. 639.

[8] Siehe *Statistisches Bundesamt*, Finanzen und Steuern, Reihe 7.1 Lohn- und Einkommensteuer 1995, S. 6 – 7.

[9] Die Erklärungsfrist endet grundsätzlich fünf Monate nach Ablauf des Kalenderjahres (§ 149 Abs. 2 AO). Gemäß § 109 AO ist eine Fristverlängerung möglich. In der Regel wird die Frist auf Grundlage eines koordinierten Ländererlasses für Steuererklärungen, die durch Personen oder Gesellschaften im Sinne des § 3 des Steuerberatungsgesetzes (StBerG) oder durch Buchstellen von Körperschaften und Vereinigungen im Sinne des § 4 Nr. 3 und 8 StBerG angefertigt werden, um vier Monate verlängert. Die Finanzämter können die Abgabefrist in einem vereinfachenden Verfahren um weitere fünf Monate verlängern.

[10] Die Festsetzungsfrist für die Einkommensteuer beträgt gemäß § 169 Abs. 2 Satz 1 Nr. 2 AO vier Jahre. Die Festsetzungsfrist beträgt zehn Jahre, soweit eine Steuer hinterzogen, und fünf Jahre, soweit sie leichtfertig verkürzt worden ist (§ 169 Abs. 2 Satz 2 AO).

[11] So auch *Rosinus*, WiSta 1991, S. 466 (467), und *Merz/Zwick*, WiSta 2001, S. 513 (516).

[12] Vgl. *Statistisches Bundesamt*, Finanzen und Steuern, Reihe 7.1 Lohn- und Einkommensteuer 1995, S. 7.

[13] Siehe *Fuß* (Fn. 6), S. 50, und *von der Lippe*, SteuerStud 1997, S. 266.

[14] *Rosinus*, WiSta 2000, S. 456 (457).

[15] So auch *Rosinus*, WiSta 1991, S. 466 (467), und *Merz/Zwick*, WiSta 2001, S. 513 (516).

[16] So *Zwick*, WiSta 1998, S. 566.

[17] Zur Thematik des Timelag zwischen festgesetzter Steuerschuld eines Veranlagungszeitraumes und kassenmäßigem Steueraufkommen siehe ausführlich bei *Löbbe/Roth*, Methoden der mittelfristigen Steuervorausschätzung, 1971, S. 46, 59 – 61, 64 und 69 – 70, *Koths*, Eine Analyse des Steueraufkommens mit Hilfe der Teilsteuerrechnung, 1978, S. 165 – 170, *Körner*, Methoden zur mehrjährigen Vorausschätzung des Einkommensteueraufkommens; ifo-Studien zur Finanzpolitik Nr. 25,

pflichtigen durchschnittlich beanspruchte Zeit bis zur Abgabe der Steuererklärung sowie die durchschnittliche Bearbeitungsdauer durch die Finanzverwaltung unterliegen Schwankungen. In Ermangelung verlässlicher Daten gestaltet sich die Zusammenführung der Steuerstatistik und des kassenmäßigen Steueraufkommens schwierig. Die Schätzung von Timelagfunktionen[18] ist in hohem Maße von Plausibilitätsüberlegungen und Erfahrungswerten abhängig, die zum einen auf nicht ausreichend repräsentative Untersuchungen zurückgehen und bei denen zum anderen die künftige Gültigkeit unsicher ist.[19]

Die Steuerstatistik ist in ihrer Konzeption als Totalerhebung angelegt. Die Einkommensteuerstatistik erfasst jedoch nur Steuerpflichtige, deren Besteuerungsunterlagen im Rahmen des Lohnsteuerjahresausgleichs[20] oder der Veranlagung zur Einkommensteuer bei den Finanzämtern eingereicht wurden, sowie nichtveranlagte Steuerpflichtige, deren Lohnsteuerkarten auf Grund der bestehenden gesetzlichen Abgabepflicht den Finanzämtern zugegangen sind.[21] Dabei werden in der Einkommensteuerstatistik nach § 26b EStG zusammenveranlagte Ehegatten grundsätzlich als ein Steuerpflichtiger ausgewiesen. Nichtveranlagte Lohnsteuerpflichtige, deren Lohnsteuerkarten nicht den Finanzbehörden zugeführt wurden,[22] sonstige nichtveranlagte Steuern vom Ertrag – insbesondere die Kapitalertragsteuer der nichtveranlagten Steuerpflichtigen – sowie die Lohnsteuer, die im Rahmen des pauschalierten Lohnsteuerabzugs abgeführt wurde,[23] erscheinen nicht in der amtlichen Steuerstatistik.

In der Steuerstatistik wird eine Vielzahl von Daten aus den Lohnsteuerkarten, Steuererklärungen und Steuerbescheiden dokumentiert. So beinhaltet die Einkommensteuerstatistik für den Veranlagungszeitraum 1995 in ca. 30 Mio. Datensätzen mit bis zu 450 Merkmalen Angaben von rund 45 Mio. Personen.[24] Diese Datensätze enthalten neben veranlagungstechnischen und sozioökonomischen Angaben insbesondere Daten zur Ermittlung der steuerlichen Bemessungsgrundlage und der Steuerschuld. Entsprechend dem bei der Veranlagung durchgeführten Berechnungsverlauf – beginnend bei den einzelnen steuerlich zu unterscheidenden Einkunftsarten[25] – sind im Wesentlichen alle Hinzurechnungen und Abzüge – wie

---

1980, S. 3 – 10, *Körner*, Probleme der Steuerschätzung, in: Staatsfinanzierung im Wandel, Hrsg. *Hansmeyer*, 1983, S. 215 (224 – 230), und *Flascha*, Probleme und Methoden der Steueraufkommensschätzung, 1985, S. 60 – 70.

[18] Schätzungen von Timelagfunktionen zwischen Steuerentstehung und Steuerzahlung finden sich bei *Löbbe/Roth* (Fn. 17), *Kröger*, Income taxes and Income Distribution – An Empirical Analysis for the FRG 1960 – 1974; SYSIFO-Research Memorandum No. 6, 1976, S. 28 – 32, *Koths* (Fn. 15), S. 165 – 170, *Körner* (Fn. 15), S. 10 – 51, und *Körner* (Fn. 15), S. 224 – 230.

[19] Da der zeitliche Ablauf des Veranlagungsverfahrens auch von der Situation und dem Verhalten der Steuerpflichtigen sowie der personellen und materiellen Ausstattung der Finanzverwaltung abhängig ist, kann der Timelag nur schwer theoretisch bestimmt oder aus der Vergangenheit abgeleitet werden. Siehe dazu auch *Härtel*, Steuerschätzung, in: Handwörterbuch der Wirtschaftswissenschaft, Bd. 7, Hrsg. *Albers* et al., 1977, S. 399 (403), *Koths* (Fn. 17), S. 168, *Körner* (Fn. 17), S. 3 – 7, und *Flascha* (Fn. 17), S. 63 – 66.

[20] Der behördliche Lohnsteuerjahresausgleich wurde 1992 durch eine Antragsveranlagung ersetzt.

[21] Zum Umfang der in die Einkommensteuerstatistik einbezogenen Steuerpflichtigen siehe auch *Statistisches Bundesamt*, Finanzen und Steuern, Reihe 7.1 Lohn- und Einkommensteuer 1995, S. 6 – 7. Für die Lohn- und Einkommensteuerpflichtigen, bei denen eine maschinelle Zwangs– oder Antragsveranlagung durchgeführt wurde, übermitteln die Finanzverwaltungen der Länder die bundeseinheitlich aufgebauten Datensätze auf maschinellen Datenträgern. Die Steuerpflichtigen aus der so genannten personellen Einkommensteuerveranlagung, die von der Finanzverwaltung nicht maschinell erfasst und gespeichert wurden, sind in der Statistik nicht enthalten (vgl. *Statistisches Bundesamt*, Finanzen und Steuern, Reihe 7.1 Einkommensteuer 1989, S. 10). In der Einkommensteuerstatistik 1986 betrug ihr Anteil einschließlich der Steuerpflichtigen, die nach Ende der Erhebungsfrist noch nicht maschinell veranlagt waren, nur noch weniger als 2% der Grundgesamtheit (vgl. *Rosinus*, WiSta 1991, S. 466).

[22] Auch wenn gemäß R 135 Abs. 13 LStR Steuerpflichtige, die am Schluss des Kalenderjahres im Besitz ihrer Lohnsteuerkarte sind und diese nicht für eine Veranlagung benötigen, sie spätestens bis zum Ablauf des Folgejahres beim Finanzamt abzugeben haben, wird nur ein Teil der nicht im Rahmen einer Steuererklärung verwendeten Lohnsteuerkarten bei den Finanzämtern tatsächlich eingereicht (vgl. auch *Statistisches Bundesamt*, Finanzen und Steuern, Reihe 7.1 Lohn- und Einkommensteuer 1995, S. 7). De shalb kommt es in der Lohnsteuerstatistik insoweit zu einer Untererfassung.

[23] Bei der Pauschalbesteuerung i.S.d. §§ 40 bis 40b EStG wird die Lohnsteuer vom Arbeitgeber übernommen. Dies ist z.B. bei Steuerpflichtigen, die nur kurzzeitig beschäftigt werden, oder bei einer Beschäftigung mit geringem Umfang und gegen geringen Arbeitslohn der Fall (§ 40a EStG). Im Besteuerungsverfahren der Pauschalisierung der Lohnsteuer gilt die Steuerschuld des Arbeitnehmers für den pauschal besteuerten Lohn als abgegolten, und die entsprechenden Einkünfte sowie die pauschale Lohnsteuer bleiben bei einer Veranlagung zur Einkommensteuer und beim Lohnsteuerjahresausgleich außer Ansatz (§ 40 Abs. 3 EStG).

[24] Rund 38 Mio. Personen erzielten Einkünfte. Vgl. *Merz/Zwick*, WiSta 2002, S. 729 (731). Zum Aufbau der Datensätze der Lohn- und Einkommensteuerstatistik 1992 siehe *Zwick*, Die erweiterte Datenverfügbarkeit nach der Modifizierung des Gesetzes über Steuerstatistiken; in: Einkommen und Vermögen in Deutschland, Hrsg. *Statistisches Bundesamt*, 1998, S. 258 – 264.

[25] Die bei der Ermittlung der einkommensteuerlichen Bemessungsgrundlage zu unterscheidenden Einkünfte ergeben sich aus § 2 Abs. 1 EStG. In der Einkommensteuerstatistik werden einige steuerliche Einkunftsarten noch differenzierter ausgewiesen; z.B. nach Rechtsform und Gewinnermittlung. Eine Übersicht über die aus der Einkünfteermittlung verfügbaren Daten findet sich bei *Merz/Zwick*, WiSta 2001, S. 513 (517 – 518).

z.B. Sonderausgaben und Aufwendungen für außergewöhnliche Belastungen – detailliert bis zur Ermittlung der steuerlichen Bemessungsgrundlage „zu versteuerndes Einkommen“[26] erfasst. Auch die gemäß § 32b EStG dem Progressionsvorbehalt unterliegenden Einkünfte und die für die besonderen Steuersätze nach §§ 34 und 34b EStG relevanten außerordentlichen Einkünfte werden in der Einkommensteuerstatistik ausgewiesen.

Ein erhebliches Informationsdefizit entsteht jedoch durch den grundsätzlich erst nach dem Abschluss der Einkünfteermittlung beginnenden Nachweis der Berechnung der steuerlichen Bemessungsgrundlage. Dies führt zu einer nicht unerheblichen Untererfassung, da z.B. die Steuerpflichtigen, deren Einnahmeüberschüsse aus Kapitalvermögen den Sparerfreibetrag nach § 20 Abs. 4 EStG nicht übersteigen, mit ihren Einnahmen aus Kapitalvermögen in der Steuerstatistik nicht erscheinen.[27] Des Weiteren fehlen wertvolle Informationen über die Höhe und die Struktur der Erträge und Aufwendungen bzw. der Einnahmen und Werbungskosten. Mit Ausnahme der Einkünfte aus nichtselbstständiger Arbeit, bei denen der Bruttolohn, die Versorgungsbezüge, der Versorgungsfreibetrag und die Werbungskosten separat in der Steuerstatistik ausgewiesen werden, liegen für die übrigen Einkunftsarten kaum Angaben über die Einkünfteermittlung vor.[28] Für viele Sachverhalte, deren Aufkommenswirkungen von besonderem steuerpolitischem Interesse sind – hier wäre z.B. an die nach § 3 EStG steuerfreien Einnahmen, die Höhe und Struktur der Abschreibungen oder die Werbungskosten bei Einkünften aus Vermietung und Verpachtung zu denken –, liegen somit aus der Veranlagung keine oder nur unvollständige Daten vor. Informationen über den Umfang der Inanspruchnahme der verschiedenen Abschreibungsregelungen können lediglich aus den in der Steuerstatistik enthaltenen Angaben aus der Anlage St. zur Einkommensteuererklärung entnommen werden. Diese Anlage ist zwar Bestandteil der Steuererklärung und ist zusammen mit den übrigen Erklärungsvordrucken abzugeben, jedoch sind die in der Anlage St. vom Steuerpflichtigen deklarierten Angaben nicht Grundlage der Veranlagung. Sie sind lediglich ergänzende Angaben für Zwecke der Steuerstatistik, die gemäß § 150 Abs. 5 AO im Rahmen der Steuererklärung erhoben werden. In der Vergangenheit haben die Finanzämter die Abgabe der Anlage St. nicht immer konsequent durchgesetzt, so dass keine vollständige Erfassung der Steuerpflichtigen erfolgte.[29] Fraglich ist auch, ob die in der Anlage St. deklarierten Angaben der Steuerpflichtigen von den Finanzämtern in ausreichendem Maße überprüft und mit den Veranlagungsergebnissen abgeglichen wurden. Vor diesem Hintergrund sind die Auswertungen, die auf den aus der Anlage St. gewonnenen Daten basieren, hinsichtlich der Vollständigkeit und Richtigkeit eingeschränkt.

Neben der Ermittlung der Bemessungsgrundlage wird in der Einkommensteuerstatistik auch die Berechnung der Steuerschuld dokumentiert. Ausgehend von der tariflichen Einkommensteuer, die sich gemäß § 32a Abs. 1 EStG aus dem Einkommensteuertarif unter Berücksichtigung der besonderen Tarifvorschriften wie z.B. des Progressionsvorbehalts (§ 32b EStG) und den besonderen Steuersätzen nach §§ 34 und

[26] Zur Ermittlung des zu versteuernden Einkommens siehe § 2 EStG; das Berechnungsschema findet sich in R 3 Abs. 1 EStR.
[27] Siehe *Merz/Zwick*, WiSta 2001, S. 513 (517).
[28] In der Einkommensteuerstatistik für den Veranlagungszeitraum 1995 wird lediglich für Gewinne aus der Veräußerung eines Betriebs, Teilbetriebs, Mitunternehmeranteils oder einer Beteiligung i.S.d. § 17 EStG die Höhe des gewährten Freibetrags dokumentiert. Der Datenkatalog der bisher noch nicht veröffentlichten Einkommensteuerstatistik für den Veranlagungszeitraum 1998 wurde diesbezüglich erweitert. So werden in der Einkommensteuerstatistik 1998 z.B. auch die Einnahmen und Werbungskosten aus Kapitalvermögen und der in Anspruch genommene Sparerfreibetrag ausgewiesen.
[29] Vgl. *Rosinus*, WiSta 2000, S. 456 (458), und *Merz/Zwick*, WiSta 2001, S. 513 (518). Es ist jedoch anzumerken, dass sich für den Veranlagungszeitraum 1995 die Erfassungsquote der Anlage St im Vergleich zu 1992 erhöht hat.

34b EStG ergibt, werden die verschiedenen Steuerermäßigungen[30] und -erhöhungen[31] ausgewiesen. Weiterhin sind in der Einkommensteuerstatistik die gemäß § 36 Abs. 2 EStG auf die festgesetzte Einkommensteuer anzurechnenden Steuern, wie die Lohnsteuer, die Kapitalertragsteuer und die Körperschaftsteuer, sowie die sich nach der Anrechnung ergebenden Steuererstattungen und -nachzahlungen aufgeführt.

Somit können aus der Einkommensteuerstatistik für den jeweiligen Veranlagungszeitraum detaillierte Angaben von allen erfassten Steuerpflichtigen gewonnen werden, die das Veranlagungsergebnis – bei den Einkünften beginnend bis zur Steuerschuld – dokumentieren. Eine intertemporale Verknüpfung der Daten der einzelnen Steuerpflichtigen im Sinne einer Längsschnittanalyse ist auf Grund der Rechtslage derzeit jedoch nicht möglich.

Das Statistische Bundesamt veröffentlicht einen Teil der in der Einkommensteuerstatistik enthaltenen Daten in Tabellen zusammengefasst in der Fachserie 14 „Finanzen und Steuern“. Die Tabellen sind vertikal in der Regel nach Größenklassen des Gesamtbetrags der Einkünfte[32] gegliedert, wobei in der Einkommensteuerstatistik 19 Klassen unterschieden werden. Für die Darstellung einiger Sachverhalte werden abweichende Größenklassenmerkmale verwendet. So werden z.B. zusammenfassende Übersichten der Lohn- und Einkommensteuerpflichtigen nach der steuerlichen Bemessungsgrundlage „zu versteuerndes Einkommen“ gegliedert. Die Tabellen enthalten die aggregierten Merkmalsausprägungen aus den Einzeldatensätzen der Steuerpflichtigen oder bestimmter Gruppen von Steuerpflichtigen – wie z.B. der nach dem Grund- oder Splittingtarif Besteuerten. Dabei werden für das jeweilige Merkmal je Größenklasse die Anzahl der Steuerpflichtigen und, soweit es sich um ein in einem Geldbetrag auszudrückendes Merkmal handelt, der Betrag in TDM bzw. T€ ausgewiesen.

In den veröffentlichten Tabellen der Steuerstatistik können somit für die jeweilige Gruppe von Steuerpflichtigen einer Größenklasse die Ermittlung der Steuerbemessungsgrundlage und die festgesetzte Steuer nachvollzogen werden. Durch die Aggregation der Daten in Klassen ergibt sich jedoch gegenüber den Einzeldatensätzen ein erheblicher Informationsverlust, da sich für jedes Merkmal nur noch je Klasse die gesamte und die durchschnittliche Ausprägung ermitteln lassen.

III. Modelle zur Simulation der Besteuerung des Einkommens

Im Folgenden werden kurz die prinzipiell für eine Steuersimulation anwendbaren Modelltypen beschrieben und einige der bisher in der Bundesrepublik Deutschland entwickelten Einkommensteuer-Simulationsmodelle vorgestellt. Da das umfangreiche Schrifttum zu dieser Thematik im Rahmen des vorliegenden Beitrags nicht abschließend gewürdigt werden kann, sei zur Vertiefung der Problematik der Steuersimulation auf die ausführliche Literatur verwiesen.[33]

[30] Vgl. dazu Teil V des EStG (Steuerermäßigungen) mit den §§ 34c bis 35 EStG.

[31] Eine Steuererhöhung kann sich u.a. durch eine Nachversteuerung gemäß §§ 30 und 31 EStDV ergeben.

[32] Der Gesamtbetrag der Einkünfte ergibt sich gemäß § 2 Abs. 3 Satz 1 EStG aus der Summe der Einkünfte, vermindert um den Altersentlastungsbetrag (§ 24a EStG) und nach Abzug des Freibetrags für Land- und Forstwirte (§ 13 Abs. 3 EStG). Vgl. dazu auch R 3 Abs. 1 EStR.

[33] Einen guten Überblick und einen Einstieg bieten z.B. *Merz*, Microsimulation – A Survey of Principles, Developments and Applications with Focus on the Static Dase and the Static Sfb 3-Microsimulation Model as an Example; Sfb 3-Arbeitspapier Nr. 268, Universität Frankfurt und Universität Mannheim 1988, *Petersen*, Simulationsmodelle für die Steuer- und Sozialpolitik; in: Wirkungsanalysen alternativer Steuer- und Transfersysteme, Hrsg. *Petersen* et al., 1990, S. 13 – 22, *Spahn* et al. (Fn. 4), und *Bork* (Fn. 4).

„Unter dem Begriff der Simulation wird in den Wirtschafts- und Sozialwissenschaften allgemein eine Technik verstanden, die darauf abzielt, die Eigenschaften und das Verhalten eines realen Systems und seiner Subsysteme auf der Grundlage eines der Realität nachgebildeten operablen Modells kennen zu lernen, indem man das Verhalten des Modells unter unterschiedlichen Bedingungen untersucht.[34] Die Simulation wird auch als Technik zur Durchführung von Experimenten an formalen Modellen bezeichnet.[35] Sie stellt in den Wirtschafts- und Sozialwissenschaften oft die einzige Alternative zum physischen Experiment dar, wobei jedoch die bei einer Simulation erzielten Ergebnisse nicht ohne weiteres verallgemeinert werden können, da sie zunächst nur für die bei der Simulation zu Grunde gelegten Prämissen gelten.

Bezogen auf das wichtige Unterscheidungsmerkmal des Aggregationsgrads der verwendeten Daten können in den Wirtschafts- und Sozialwissenschaften grundsätzlich drei Arten von Simulationsmodellen unterschieden werden:[36]

- Modelle, die im Wesentlichen auf den Aggregaten der volkswirtschaftlichen Gesamtrechnung basieren, wie die makroökonomischen Modelle und die allgemeinen Gleichgewichtsmodelle (hohes Aggregationsniveau)[37]
- Gruppenmodelle, die an ausgewählten Merkmalen für homogene Gruppen von Wirtschaftssubjekten ansetzen (mittleres Aggregationsniveau), und
- mikroanalytische Modelle, die auf der Modellierung einzelner Mikroeinheiten basieren (starke Disaggregation)[38].

Makroökonomische Modelle und Gleichgewichtsmodelle sind für die Analyse des Aufkommens der Einkommensteuer nur eingeschränkt geeignet, da sie – bedingt durch den hohen Aggregationsgrad – individuelle Merkmale der Steuerpflichtigen und strukturelle Faktoren weder im Modellansatz noch in den Ergebnissen ausreichend berücksichtigen.[39]

Die stärker disaggregierten Gruppenmodelle und die mikroanalytischen Modellansätze weisen diesbezüglich in ihrem Aufbau Vorteile auf. Da die mikroökonomischen Modellansätze auch individuelle Merkmale der Steuerpflichtigen explizit erfassen und modellieren und daher die Ermittlung der Bemessungsgrundla-

[34] *Spahn* et al. (Fn. 4), S. 6. So ähnlich auch *Spahn*, Die Besteuerung der persönlichen Einkommen in der BRD – System und Modell, 1972, S. 109, und *Merz*, Statische mikroanalytische Simulation – Anforderungsprofil und Lösungsmöglichkeiten; Sfb 3-Arbeitspapier Nr. 124, Universität Frankfurt und Universität Mannheim 1983, S. 2. Zur Simulation sozio-ökonomischer Systeme siehe *Pagenkopf*, Simulation; in: Handwörterbuch der Wirtschaftswissenschaft, Bd. 6, Hrsg. *Albers* et al., 1981, S. 536 – 549.

[35] Vgl. *Pagenkopf* (Fn. 34), S. 539.

[36] Ausführlich dazu *Spahn* et al. (Fn. 4), S. 6 – 11, und *Bork* (Fn. 4), S. 63 – 72. Neben der Unterscheidung bezüglich des Dissaggregationsgrades und des sachlichen Untersuchungsgegenstandes werden die Simulationsmodelle u.a. auch in dynamische und statische unterteilt, d.h., sie werden danach unterschieden, wie die individuellen Biographien der Untersuchungseinheiten fortgeschrieben werden. Ein weiteres Unterscheidungskriterium stellt die Berücksichtigung von Verhaltensreaktionen dar. Vgl. dazu u.a. *Krupp/Wagner*, Vierteljahreshefte zur Wirtschaftsforschung 1982, S. 5 (11 – 13), *Petersen*, Mikroökonomische Simulationsmodelle zur Erhöhung der Rationalität; in: Finanz- und Sozialpolitik, Finanzwissenschaftliche Arbeitspapiere Nr.8, Justus-Liebig-Universität Gießen 1986, S. 6 – 8, *Merz* (Fn. 33), S. 4 – 9, und *Bork* (Fn. 4), S. 72 – 84.

[37] Vgl. dazu *Teichmann*, Vierteljahreshefte zur Wirtschaftsforschung, 1974, S. 228, *Bös/Genser*, Steuerfunktionen in Prognose- und Entscheidungsmodellen, 1977, *Pauly*, Ökonometrische Analyse der Einkommensbesteuerung, 1983, *Spahn* et al. (Fn. 4), S. 7, *Bork* (Fn. 4), S. 63 – 66. Einen Überblick geben z.B. *Hujer/Hansen/Klein*, Der Beitrag makroökonomischer Modelle zur Beurteilung steuerpolitischer Maßnahmen, Sfb3-Arbeitspapier Nr.273, J. W. Goethe-Universität Frankfurt und Universität Mannheim, 1988, S. 3.

[38] Der erste sozioökonomische Mikrosimulationsansatz geht auf *Orcutt* zurück (*Orcutt*, The Review of Economics and Statistics 1957, S. 116). Zu den Grundlagen und der Anwendung mikroanalytischer Modelle siehe u.a. *Krupp/Wagner*, Vierteljahreshefte zur Wirtschaftsforschung 1982, S. 5, und *Orcutt/Merz/Quinke* (Hrsg.), Microanalytic Simulation Models to Support Social and Financial Policy, 1986. Ausführlich zur Konstruktion mikroanalytischer Simulationsmodelle siehe *Vetterle*, Konstruktion und Simulation mikroanalytischer Modelle – Die Methode der Mikrosimulation und ihre Anwendung, 1982.

[39] So auch *Galler*, Microsimulation of Tax-Transfer Systems; in: Prospects and Limits of Simulation Models in Tax and Transfer Policy, Hrsg. *Brunner/Petersen*, 1990, S. 279 – 300, S. 280, *Spahn* et al. (Fn. 4), S. 7, und *Bork* (Fn. 4), S. 66.

ge und der Steuer präziser simuliert werden kann, sind genauere und differenziertere Aussagen zum Steueraufkommen möglich.

Im Allgemeinen besitzen Gruppenmodelle – bedingt durch ihren geringeren Modellumfang – im Vergleich zu den mikroanalytischen Modellen eine relativ einfache und transparente Struktur. Dies erleichtert ihre Implementierung und Modifikation und macht sie zu einem flexiblen und kostengünstigen Analyseinstrument zur Untersuchung von Aufkommenswirkungen. Aber insbesondere der Informationsverlust auf Grund der verwendeten aggregierten Daten und die begrenzte Möglichkeit einer flexiblen Merkmalsgliederung schränken den Anwendungsbereich von Gruppenmodellen ein.[40] Bei vielen Fragestellungen der Steueraufkommensanalyse sind diese Einschränkungen jedoch vertretbar. Ein hoher Disaggregationsgrad, der nur durch mikroanalytische Simulationsmodelle erreicht werden kann, ist insbesondere für Verteilungsanalysen und Verhaltenssimulationen sowie für eine umfassende, mehrere Steuer- und Transfersysteme einbeziehende Simulation notwendig.

Bei der mikroanalytischen Simulation[41] wird jede einzelne Mikroeinheit mit ihren Merkmalen unmittelbar einbezogen. Dies kann anhand typisierter Einzelfälle, einer Stichprobe oder der Grundgesamtheit erfolgen. Der Vorteil von umfassenden Strukturinformationen kann aber nur genutzt werden, wenn eine entsprechende Vielzahl von Merkmalen der Mikroeinheiten als Datenbasis zur Verfügung steht. Um eine möglichst realitätsnahe Simulation zu erreichen, müssen Abhängigkeiten und Wirkungen im Modell implementiert werden, was dazu führt, dass mit steigender Merkmalsanzahl auch die Komplexität des Modells zunimmt.

Auch wenn die mikroanalytischen Modelle aus theoretischer Sicht den Gruppenmodellen überlegen sind,[42] setzen die Probleme, die im Zusammenhang mit dem Zugang zu den notwendigen Daten, der Implementierung des Modells und der Verwaltung, Aktualisierung und Fortschreibung des Datenbestandes auftreten, ihrer Anwendung Grenzen.[43] Daher war der Einsatz von mikroanalytischen Modellen auf dem Gebiet der Steueraufkommens- und Verteilungsanalyse sowie der Steuerschätzung meist nur bei größeren Forschungsprojekten sowie bei Ministerien oder Forschungsinstituten möglich, die über die entsprechenden Daten und Ressourcen verfügen. Hier sind insbesondere das im Rahmen des Sonderforschungsbereichs 3 (Sfb3) „Mikroanalytische Grundlagen der Gesellschaftspolitik“ DFG an der Universität Frankfurt und der Universität Mannheim durchgeführte Projekt[44] zur Erstellung eines Simulationsmodells der Einkommensbesteuerung auf der Basis des sozio-ökonomischen Panels[45], das Potsdamer Mikro-

[40] Siehe dazu *Essen/Kassella*, Die Einkommensangaben im sozio-ökonomischen Panel des Sonderforschungsbereichs 3 und ihre Relevanz für steuerpolitische Simulationen; in: Aufgaben und Probleme der Einkommensteuerstatistik, Hrsg. *Reich*, S. 133 (S. 134 – 136), *Spahn* et al. (Fn. 4), S. 24 – 27, und *Bork* (Fn. 4), S. 67 – 69.

[41] Zur Methode der Mikrosimulation bei der Untersuchung steuerpolitischer Fragestellungen siehe statt vieler *Spahn* et al. (Fn. 4).

[42] Vgl. *Petersen* (Fn. 33), S. 19.

[43] Ausführlich dazu *Spahn* et al. (Fn. 4), S. 32 – 41 und S. 107 – 169, sowie Petersen (Fn. 33), S. 19 – 20. Zur Eignung von Daten zur Steuersimulation siehe auch *Krupp/Wagner*, Vierteljahreshefte zur Wirtschaftsforschung 1982, S. 5 (13 – 15), *Gyárfás*, Ein Simualtionsmodell der Einkommensbesteuerung auf der Grundlage synthetischer Mikrodaten, 1990, S. 13 – 17, *Hüther*, Zu den methodisch-empirischen Grundlagen der Simulation alternativer steuer- und sozialrechtlicher Regelungen für die Bundesrepublik Deutschland; in: Wirkungsanalysen alternativer Steuer- und Transfersysteme, Hrsg. *Petersen* et al., 1990, S. 23 (28 – 40), und *Bork/Petersen* (Fn. 2).

[44] Zum Mikrosimulationsmodell des Sfb 3 siehe statt vieler *Galler/Ott*, Das dynamische Mikrosimulationsmodell des Sonderforschungsbereichs 3; in: Mikroanalytische Grundlagen der Gesellschaftspolitik, Bd. 2, Hrsg. *Hauser* et al., 1994, S. 399 – 427. Einen Überblick über die Arbeiten des von 1979 bis 1990 durch die DFG geförderten Sfb 3 gibt u.a. *Hauser*, Die Arbeiten des Sonderforschungsbereichs 3; in: Mikroanalytische Grundlagen der Gesellschaftspolitik, Bd. 1, Hrsg. *Hauser* et al, 1994, S. 1 – 34. Zu den Modellierungskonzepten der an der Universität Frankfurt entwickelten Mikrosimulationsmodelle siehe *Galler*, Politikanalyse mit Mikrosimulationsmodellen – die Frankfurter Modelle; in: Mikroanalytische Grundlagen der Gesellschaftspolitik, Bd. 1, Hrsg. *Hauser* et al, 1994, S. 113 (115 – 120).

[45] Zum Einkommensteuersimulationsmodell vgl. u.a. *Essen/Kassella/Landua*, Ein Simualtionsmodell der Einkommensbesteuerung auf der Basis des sozio-ökonomischen Panels; Sfb3-Arbeitspapier Nr. 188, Universität Frankfurt und Universität Mannheim 1986, *Essen/Kassella*, Die Einkommensangaben im sozio-ökonomischen Panel des Sonderforschungsbereichs 3 und ihre Relevanz für steuerpolitische Simulationen; in: Aufgaben und Probleme der Einkommensteuerstatistik, Hrsg. *Reich*, 1988, S. 133 (151 – 155), und *Spahn* et al. (Fn. 4), S. 102 – 106.

simulationsmodell[46] und das Einkommensteuermodell des BMF[47] zu nennen.[48]

Simulationsmodelle wie z.B. das ehemalige Einkommensteuermodell des BMF, die auf den disaggregierten Daten der amtlichen Lohnsteuer- und Einkommensteuerstatistik basieren[49], stellen im Grunde keine Mikrosimulationsmodelle im engeren Sinne, sondern eher Gruppenmodelle dar,[50] die mit synthetischen Daten[51] operieren. Das erste in der Bundesrepublik Deutschland entwickelte Steuersimulations-Gruppenmodell ist das Modell von Spahn (1972). Bei diesem Modell wurden auf Grundlage der Daten der Steuerstatistik mittels Spline-Interpolation stetige Einkommensverteilungsfunktionen generiert, die nach Maßgabe der unterstellten Einkommensentwicklung fortgeschrieben wurden. Spahn führte mit diesem Modell u.a. Schätzungen zum Aufkommen der Lohn- und Einkommensteuer[52] und die Simulation der langfristigen Veränderung der Einkommensverteilung bei einer progressiven Einkommensteuer[53] durch.[54]

Auch bei dem von Petersen (1977) entwickelten Simulationsmodell wurden aus den gruppierten Daten der Lohnsteuerstatistik durch lineare Interpolation stetige Einkommensverteilungsfunktionen gewonnen. Einen vom Prinzip identischen, aber vergleichsweise einfachen Ansatz wählte *Boss* in seinem Lohnsteuermodell.[55] Bei dem noch heute vom Institut für Weltwirtschaft (IfW) verwendeten Modell wird die Datenbasis aus den Einkommensschichtungen der Lohnsteuerstatistik wie folgt gebildet: Jede Klasse der Einkommensschichtung repräsentiert einen Steuerfall, dem als Gewichtung die Besetzungshäufigkeit und als Einkommen das Durchschnittseinkommen der Klasse zugeordnet wird. Jedoch ergibt sich bei der Berechnung der Lohnsteuer entsprechend der Vorgehensweise von Boss im Vergleich zum tatsächlichen, in der Lohnsteuerstatistik ausgewiesenen Aufkommen ein tendenziell zu geringes Steueraufkommen,[56] da die Auswirkungen des progressiven Einkommensteuertarifs nicht adäquat simuliert werden.[57]

Die Vorgehensweise von *Boss* ist für die Bestimmung von Aufkommenswirkungen bei einfachen Fragestellungen prinzipiell ausreichend. Bei der Untersuchung von Sachverhalten, für die Häufigkeitsverteilungen stärker berücksichtigt werden müssen, findet die Anwendung des Modells jedoch seine Grenzen, da die Anzahl der „Steuerfälle" im Modell auf die Anzahl der Einkommensklassen der Steuerstatistik beschränkt

---

[46] Vgl. dazu ausführlich *Bork* (Fn. 4), S. 112 – 153.

[47] Siehe *Lietmeyer*, Das Einkommensteuermodell des Bundesministeriums der Finanzen; in: Möglichkeiten und Grenzen sozio-ökonomischer Modelle, Bd. 2, Hrsg. *Bendisch/Hoschka*, 1983, S. 7 – 26, *Lietmeyer*, Microanalytic Tax Simulation Models in Europe: Development and Experience in the German Federal Ministry of Finance; in: Microanalytic Simulation Models to Support Social and Financial Policy, Hrsg. *Orcutt* et al., 1986, S. 139 – 152, Das mikroanalytische Einkommensteuermodell des BMF – Aufbau, Anwendung und Weiterentwicklung, 1989, und *Spahn* et al. (Fn. 4), S. 93 – 102.

[48] Zur geschichtlichen Entwicklung der mikroökonomischen Simulationsmodelle siehe u.a. *Vetterle* (Fn. 38), S. 4 – 7.

[49] Vgl. *Lietmeyer* (1989 Fn. 47), S. 3 – 5 und 8 – 10, sowie *Gyárfás* (Fn. 43), S. 20 – 28 und 85 – 108. Diese synthetische Datenbasis wurde in den Jahren 1987 und 1988 im Auftrag des BMF und des BMJFFG in der Gesellschaft für Mathematik und Datenverarbeitung (GMD) entwickelt. Das BMF war jedoch bestrebt, mittelfristig eine Umstellung auf eine auf Veranlagungsdaten basierende Stichprobe durchzuführen, und wollte die dem BMF zur Verfügung stehende Geschäftsstatistik über die maschinelle Einkommensteuerveranlagung einbeziehen (vgl. *Lietmeyer* (Fn. 47), S. 3).

[50] So *Spahn* et al. (Fn. 4), S. 91.

[51] Von synthetischen Daten wird in diesem Zusammenhang gesprochen, wenn die Daten aus aggregierten Daten abgeleitet werden und im Simulationsmodell wie Mikrodaten verwendet werden (vgl. *Gyárfás* (Fn. 43), S. 18 – 19). Zur Konstruktion synthetischer Mikrodaten aus aggregierten Daten siehe *Gyárfás/Quinke*, Allgemeines Statistisches Archiv 1993, S. 149.

[52] Vgl. *Spahn* (Fn. 32) und *Spahn*, Konjunkturpolitik 1973, S. 117.

[53] Vgl. *Spahn*, Konjunkturpolitik 1973, S. 295 und *Spahn*, Public Finance 1975, S. 231.

[54] Das Modell von *Spahn* wurde seinerzeit zur Unterstützung der Arbeit des BMF-Arbeitskreises „Steuerschätzungen" geschaffen und vom DIW weiterentwickelt (vgl. *Spahn* et al. (Fn. 4), S. 91).

[55] Vgl. *Boss*, Ein Modell zur Simulation des Lohnsteueraufkommens in der Bundesrepublik Deutschland – ein Beispiel für die Nutzbarmachung sekundärstatistischer Daten; in: Wirtschafts- und Sozialstatistik – Empirische Grundlagen politischer Entscheidungen, Hrsg. *Hanau* et al., 1986, S. 326 – 344, *Boss*, A Modell for Simulating Wage Income Tax Revenues in the Federal Republic of Germany – Structure, Predicitive Power and Application; in: Simulation Models in Tax and Transfer Policy, Hrsg. *Brunner/Petersen*, 1990, S. 19 – 30, und *Boss/Elendner*, Ein Modell zur Simulation des Lohnsteueraufkommens in Deutschland; Kieler Arbeitspapier Nr.988, Institut für Weltwirtschaft, 2000.

[56] Siehe *Boss* (1986 Fn. 55), S. 337, und *Boss/Elendner* (Fn. 55), S. 10.

[57] Dazu ausführlich *Gyárfás* (Fn. 43), S. 19 und 32 – 34, der eine Obergrenze für das Ausmaß der systematischen Unterschätzung bestimmt und der folgende Abschnitt dieses Beitrags.

ist. Diese Unzulänglichkeit kann durch die Verwendung von Verteilungsfunktionen behoben werden.

Die formale Nachbildung einer empirischen Häufigkeitsverteilung aus aggregierten Daten kann grundsätzlich durch zwei methodische Ansätze erreicht werden:[58]

- die Verwendung analytischer Verteilungsfunktionen, deren Parameter aus dem empirischen Material durch Approximation bestimmt werden,[59] oder
- den Einsatz von Interpolationsfunktionen, wie sie z.B. *Spahn* (1972), *Petersen* (1977) und *Gyárfás* (1990) in ihren Modellen verwenden.

Die Pareto- und die Lognormalverteilung stellen die klassischen analytischen personellen Einkommensverteilungen dar.[60] Auf Grund ihrer Nachteile bei der Beschreibung empirischer Einkommensverteilungen entstanden neue Ansätze,[61] wobei jedoch diese Verteilungen Spezialfälle der „verallgemeinerten Beta-Verteilung der zweiten Art" sind[62]. Die „verallgemeinerte Beta-Verteilung der zweiten Art" ist die flexibelste, aber auf Grund der Parameteranzahl auch die am schwierigsten handhabbare Verteilung.[63] Die mathematische Approximation der analytischen Verteilungsfunktionen an die empirischen Verteilungen ist aufwendig und es ergeben sich i.d.R. erhebliche Abweichungen, insbesondere in den oberen und unteren Einkommensbereichen.[64] Gelingt keine akzeptable mathematische Approximation, muss auf einen theoretischen Erklärungsansatz für die empirische Einkommensverteilung verzichtet und eine Interpolation vorgenommen werden. Es ist allerdings anzumerken, dass der Nutzen der Unterstellung einer analytischen Verteilungsfunktion beschränkt ist, da eine brauchbare ökonomische Interpretation der Parameter i.d.R. nicht gegeben ist.[65]

Bei Ex-post-Steueraufkommensanalysen kann auf einen analytischen Einkommensverteilungsansatz verzichtet werden. Bei der Untersuchung, deren Ergebnisse im Folgenden vorgestellt werden, wurde bei der Gruppensimulation die Einkommensverteilung mittels linearer Interpolation bestimmt. Abweichend von der Vorgehensweise bei Spahn (1972), Petersen (1977) und Gyárfás (1990), wurde jedoch keine

---

[58] Vgl. *Spahn*, Konjunkturpolitik 1973, S. 117 (120).

[59] Eine Übersicht über unterschiedliche Ansätze der empirisch orientierten Verteilungstheorie findet sich bei *Krupp*, Empirische Ansätze zur Erklärung der personellen Einkommensverteilung; in: Neue Aspekte der Verteilungstheorie, Hrsg. *Bombach* et al., 1974, S. 99 (135 Übersicht 1). Siehe auch *Kröger* (Fn. 18), S. 6 und 23, der für die Beschreibung der Einkommensverteilung auf Basis der Lohnsteuer- und Einkommensteuerstatistiken der Jahre 1961, 1965, 1968 und 1971 die zweiparametrige Lognormalverteilung wählt. Eine umfangreiche Untersuchung hat z.B. *Körner* (Fn. 17), S. 164 – 222, durchgeführt, der die zweiparametrige Lognormalverteilung, die zweiparametrige Gammaverteilung und die dreiparametrige Lebensdauerverteilung auf die empirischen Einkommensverteilungen in der Bundesrepublik Deutschland von 1965, 1968, 1971 und 1974 anwendet. Die zweiparametrige Lognormalverteilung wurde auch im ökonometrischen Konjunkturmodell SYSIFO (vgl. *Kröger*, Direkte Steuern; in: SYSIFO – Ein ökonometrisches Konjunkturmodell für die Bundesrepublik Deutschland, Hrsg. *Hansen/Westphal*, 1983, S. 393 (397), für den unteren Einkommensbereich im Lohnsteuermodell von *Pauly* (Fn. 37) und im Lohnsteuermodell des Sfb 3 (vgl. *Klein*, Die Neuschätzung des Lohnsteueraufkommens im makroökonometrischen Modell des Sfb 3, Sfb 3-Arbeitspapier Nr.232, J. W. Goethe-Universität Frankfurt und Universität Mannheim, 1987) zur Beschreibung der Einkommensverteilungen aus den Lohnsteuerstatistiken von 1961 bis 1983 verwendet. Im oberen Einkommensbereich approximiert *Pauly* (Fn. 37) das makroökonomische Lohnsteueraufkommen mittels der dreiparametrigen Lognormalverteilung.

[60] Vgl. z.B. *Krelle*, Verteilungstheorie, 1962, S. 269 – 282, und *Blümle*, Theorie der Einkommensverteilung, 1975, S. 29 – 30. Ausführlich zur Anwendung von Pareto-Verteilungen zur Modellierung von personellen Einkommensverteilungen siehe *Ziebach*, Die Modellierung der personellen Einkommensverteilung mit verallgemeinerten Pareto-Kurven, 2000.

[61] Vgl. *Blümle* (Fn. 60), S. 29.

[62] Dies zeigten *McDonald,* Econometrica 1984, S. 647, und *McDonald/Xu*, Journal of Econometrics 1995, S.133. Vgl. auch *Kleiber*, Halbordnungen von Einkommensverteilungen, 2000, S. 30.

[63] Vgl. *Kleiber* (Fn. 62), S. 31.

[64] Zur mangelnden Güte der Anpassung in den Bereichen sehr hoher und sehr niedriger Einkommen siehe z.B. *Krupp*, Theorie der personellen Einkommensverteilung – Allgemeine Grundzüge und verteilungspolitische Simulation, 1968, S. 50 – 54, *Krupp* (Fn. 59), S. 104 – 105, *Spahn* (Fn. 34), S. 129, und *Spahn*, Konjunkturpolitik, 1973, S. 117 (120). Auch *Petersen* (Personelle Einkommenbesteuerung und Inflation – Eine theoretisch-empirische Analyse der Lohn- und veranlagten Einkommensteuer in der Bundesrepublik Deutschland, 1977) verzichtete auf die Anwendung einer analytischen Einkommensverteilung (Pareto- und Lognormalverteilung), da auch die mathematische Approximation mit Polynomen höheren Grades keine akzeptable stetige Funktion ergab. *Körner* (Fn. 17) erzielte m.E. ebenfalls keine für die Steueraufkommensanalyse brauchbaren Ergebnisse. Die Verwendung der zweiparametrigen Lognormalverteilung im Lohnsteuermodell des Sfb 3 führte – trotz der Probleme in den „Randklassen" – zu einer relativ guten Dissaggregation hinsichtlich der relativen Besetzung der Bruttolohnklassen (vgl. *Klein* (Fn. 59), S. 16 und 25 – 28). Über die Güte der Anpassung des ermittelten Gesamteinkommens und des Steueraufkommens an die empirischen Werte finden sich jedoch bei *Klein* keine Ausführungen.

[65] So *Krupp* (Fn. 59), S. 104.

stetige Funktion, sondern eine arithmetische Zahlenfolge, d.h. eine diskrete Funktion für die Einkommensverteilung gewählt. Die Generierung von diskreten Einkommensverteilungsfunktionen ist für die Steueraufkommensanalyse ausreichend, da der Definitionsbereich der Einkommensteuertariffunktion nur die Menge der natürlichen Zahlen, also diskrete Argumente, umfasst.[66] Außerdem wird, anders als bei Spahn (1972) und Petersen (1977), in deren Steuersimulationsmodellen die kumulierten Häufigkeiten der Steuerpflichtigen ohne Berücksichtigung des Klassengesamteinkommens interpoliert werden, wie im Modell von Gyárfás (1990), das aggregierte zu versteuernde Einkommen aller Steuerpflichtigen einer Klasse bei der Interpolation einbezogen.[67]

Bei dem im Folgenden vorgestellten diskreten Modell zur Simulation der persönlichen Einkommensbesteuerung auf Basis klassierter Daten der Einkommensteuerstatistik wird dadurch eine Disaggregationsgenauigkeit erreicht, die in jeder Klasse zu einer 100%-igen Anpassung an das aus der Steuerstatistik bekannte aggregierte zu versteuernde Einkommen der jeweiligen Klasse führt.

IV. Konzept eines diskreten Modells zur Simulation des Aufkommens der Einkommensteuer auf Basis klassierter Daten der Einkommensteuerstatistik

Im Folgenden wird ein diskretes Einkommensteuersimulationsmodell auf Basis klassierter Daten vorgestellt. Die Anforderungen an das zu entwickelnde Gruppenmodell bestanden darin, auf Basis der aggregierten Daten der Einkommensteuerstatistik mit den verfügbaren Ressourcen schnell und flexibel die Aufkommenswirkungen alternativer Zustände, insbesondere die finanziellen Auswirkungen bestimmter steuerlicher Regelungen ermittelt zu können. Anschließend an die Beschreibung des Gruppenmodells werden die Ergebnisse von Einzel- und der Gruppensimulationsrechnungen miteinander verglichen, um ein Bild von der Genauigkeit des Gruppenmodells zu vermitteln.

IV.1. Diskrete Einkommensverteilung

Die absolute Häufigkeit der Steuerpflichtigen mit einem bestimmten zu versteuernden Einkommen zvE ist $h_{(zvE)}$ und ergibt sich für das geschlossene Einkommensintervall i mit den Intervallgrenzen $[a_i,b_i]$ – wobei $ai_{+1}=b_i+1$ – aus der diskreten Dichtefunktion der Steuerpflichtigen:

$$(1.1) \qquad h_i = \sum_{zvE=a_i}^{b_i} h_{(zvE)}.$$

Diese Zahlenfolge ist eine eindeutige Abbildung einer Menge von natürlichen Zahlen (zu versteuerndes Einkommen) auf eine Menge von ganzen Zahlen (absolute Häufigkeit der Steuerpflichtigen).

Die Summe der zu versteuernden Einkommen der Steuerpflichtigen im Intervall i ist zvE_i und lässt sich aus der Dichtefunktion wie folgt bestimmen:

$$(1.2) \qquad zvE_i = \sum_{zvE=a_i}^{b_i} h_{(zvE)}\, zvE.$$

[66] Gemäß § 32 Abs. 2 EStG ist der Einkommensteuertarif nur auf volle DM bzw. Euro-Beträge anzuwenden.
[67] Siehe dazu *Gyárfás* (Fn. 43), S. 30 – 31.

Wird auf die steuerliche Bemessungsgrundlage zvE der Einkommensteuertarif nach § 32a EStG angewendet, ergibt sich vorbehaltlich der besonderen Tarifvorschriften[68], der Steuerermäßigungen und -erhöhungen sowie der nach § 36 Abs. 2 EStG anzurechnenden Steuern die Einkommensteuer $t_{(zvE)}$. Die Summe der nach § 32a EStG ermittelten Einkommensteuer aller Steuerpflichtigen des Intervalls i ist T_i und ergibt sich aus:

(1.3) $$T_i = \sum_{zvE=a_i}^{b_i} h_{(zve)} t_{(zvE)}.$$

IV.2. Daten nach Größenklassen des zu versteuernden Einkommens

Wie bereits ausgeführt, enthalten die Tabellen der veröffentlichten Einkommensteuerstatistik – getrennt nach Steuerpflichtigen, die mit dem Grundtarif, und nach Steuerpflichtigen, die mit dem Splittingtarif besteuert werden – aggregierte Daten für eine Vielzahl von steuerrelevanten Sachverhalten. Dabei werden je Klasse die Anzahl der Steuerpflichtigen und die Ausprägung des Sachverhalts in einem Geldbetrag ausgewiesen. Für Simulationsrechnungen zum Zwecke der Steueraufkommensanalyse eignen sich bei Anwendung eines Gruppenmodells am besten Daten, die nach Größenklassen des zu versteuernden Einkommens gruppiert sind, da der Wertebereich für die steuerliche Bemessungsgrundlage der Steuerpflichtigen einer Klasse durch die Klassengrenzen unmittelbar vorgegeben und somit die Interpolation der Verteilung der Steuerpflichtigen auf dieses Intervall begrenzt wird.

Aus den nach Größenklassen des zu versteuernden Einkommens gegliederten Daten der Einkommensteuerstatistik sind u.a. für jede Klasse die Anzahl der Steuerpflichtigen, für die ein zu versteuerndes Einkommen festgestellt wurde, die Summe der zu versteuernden Einkommen dieser Steuerpflichtigen und die Summe der festgesetzten Einkommensteuer zu entnehmen. Diese Datenbasis kann – jeweils für die Steuerpflichtigen, die mit dem Grundtarif, und für die Steuerpflichtigen, die mit dem Splittingtarif besteuert werden – formal wie folgt dargestellt werden:

Gegeben ist eine Klasseneinteilung nach dem Merkmal „zu versteuerndes Einkommen“ *zvE* für $i=1$ bis n Klassen mit den Klassengrenzen $[a_i, b_i]$, wobei $a_1=-\infty$, $b_1=0$, $a_2=1$ und $b_n=\infty$. Für jede Klasse i ist u.a. Folgendes bekannt:

- die Klassenhäufigkeit h_i (Anzahl der Steuerpflichtigen der Klasse, für die ein zu versteuerndes Einkommen festgestellt wurde),
- die Summe der zu versteuernden Einkommen zvE_i der Steuerpflichtigen der Klasse und
- die Summe der festgesetzten Einkommensteuer ESt_i der Steuerpflichtigen der Klasse.

Die festgesetzte Einkommensteuer aller Steuerpflichtigen ESt ergibt sich aus der Anwendung aller relevanten Tarifvorschriften und der Berücksichtigung der Steuerermäßigungen und -erhöhungen – jedoch vor den nach § 36 Abs. 2 EStG anzurechnenden Steuern.[69]

[68] Eine von § 32a EStG abweichende Einkommensteuer kann sich z.B. durch die Anwendung des Progressionsvorbehalts gemäß § 32b EStG oder der besonderen Steuersätze für außerordentliche Einkünfte gemäß der §§ 34 und 34b EStG ergeben.
[69] Vgl. § 2 Abs. 6 Satz 1 EStG.

In der Einkommensteuerstatistik wird zwar auch die Summe der tariflichen Einkommensteuer der Klasse, die sich nach der Anwendung der besonderen Tarifvorschriften und vor Berücksichtigung der Steuerermäßigungen und -erhöhungen ergibt, jedoch nicht T_i ausgewiesen. Den aggregierten Daten der Einkommensteuerstatistik sind zudem nicht die absolute Häufigkeit der Steuerpflichtigen mit einem bestimmten zu versteuernden Einkommen $h_{(zvE)}$, die Summe dieser zu versteuernden Einkommen $h_{(zvE)}$zvE sowie die entsprechende Einkommensteuer nach § 32a EStG $h_{(zvE)}t_{(zvE)}$ zu entnehmen. Aus der Summe der zu versteuernden Einkommen und der Anzahl der Steuerpflichtigen der Klasse lässt sich lediglich das durchschnittliche zu versteuernde Einkommen der Klasse $\overline{zvE_i} = \frac{zvE_i}{h_i}$ bestimmen. Weitere Informationen, aus denen auf die Verteilung der Steuerpflichtigen in der Klasse geschlossen werden kann, liegen nicht vor. Da sich bei der Veranlagung die festgesetzte Einkommensteuer ESt nach der Anwendung aller relevanten Tarifvorschriften – also auch des Progressionsvorbehalts gemäß § 32b EStG oder der besonderen Steuersätze gemäß § 34 EStG sowie der Steuerermäßigungen und -erhöhungen – ergibt, können mittels der aus der Einkommensteuerstatistik bekannten Summen der festgesetzten Einkommensteuer der jeweiligen Einkommensklassen (ESt_i) keine zusätzlichen Informationen über die Verteilung der Steuerpflichtigen gewonnen werden.[70] Bei identischen Bemessungsgrundlagen kann dadurch, dass sich bei jedem Steuerpflichtigen die Tarifvorschriften sowie die Steuerermäßigungen und -erhöhungen individuell unterschiedlich auswirken, eine verschieden hohe Einkommensteuer festgesetzt werden, so dass zwischen der Bemessungsgrundlage „zu versteuerndes Einkommen“ und der festgesetzten Einkommensteuer keine strenge funktionale Beziehung besteht.

Aus dem beschriebenen aggregierten Datenmaterial der Einkommensteuerstatistik sollen mit Hilfe eines diskreten Gruppensimulationsmodells die finanziellen Auswirkungen für alternative Zustände berechnet werden.

Bei einer Simulationsrechnung auf Basis von klassierten Daten tritt durch den progressiven Einkommensteuertarif u.a. folgendes, bereits in Abschnitt III. erwähntes, Problem auf. Wird das Aufkommen der Einkommensteuer wie im Lohnsteuermodell von Boss[71] mit Hilfe des durchschnittlichen zu versteuernden Einkommens der Klassen bestimmt, indem für jede Klasse die Einkommensteuer auf das durchschnittliche zu versteuernde Einkommen der Klasse $t_{(\overline{zvE_i})}$ mit der Anzahl der Steuerpflichtigen der Klasse hin multipliziert wird, ergibt sich grundsätzlich ein zu geringes Steueraufkommen. Dies ist zum einen darin begründet, dass im Bereich des progressiven Anstiegs des Einkommensteuertarifs die Einkommensteuer auf die durchschnittliche Bemessungsgrundlage einer Klasse die Progressionswirkung innerhalb der Klasse nicht vollständig abbildet. Zum anderen lassen sich die Auswirkungen eines Übergangs zwischen zwei Tarifbereichen des Steuertarifs innerhalb einer Klasse nicht nachbilden, da das durchschnittliche zu versteuernde Einkommen der Klasse nur in einem Tarifbereich liegen kann.[72] Dies wirkt sich insbesondere bei der

[70] Dies gilt auch für das in der Einkommensteuerstatistik ausgewiesene Merkmal „tarifliche Einkommensteuer“, da die tarifliche Einkommensteuer ebenfalls unter Anwendung der besonderen Tarifvorschriften ermittelt wird.
[71] Vgl. Fn. 55.
[72] Hier ist insbesondere der Übergang von der Nullzone des Steuertarifs, die durch den Grundfreibetrag bestimmt wird, problematisch, da in dem Fall eines unterhalb des Grundfreibetrags liegenden durchschnittlichen zu versteuernden Einkommens der Klasse die aggregierte Einkommensteuer der Klasse null betragen würde.

Simulation der Aufkommenswirkungen von modifizierten Bemessungsgrundlagen und Steuertarifen mit unterschiedlichen Tarifbereichen aus.

Um bei der Bestimmung des Aufkommens der Einkommensteuer mittels einer Simulationsrechnung auf Grundlage von klassierten Daten diese Ungenauigkeiten zu verringern, wird im Folgenden unter Anwendung der linearen Interpolation ein einfaches diskretes Modell für die Verteilung der Steuerpflichtigen innerhalb einer Klasse entwickelt. Die lineare Interpolation fordert, m Glieder zwischen zwei Zahlen z_1 und z_2 mit der Differenz $z_2 - z_1 = d$ so einzubeschreiben, dass eine endliche arithmetische Zahlenreihe entsteht, deren erstes Glied z1 und deren (m+2)-tes Glied z2 ist. Nennt man d die Differenz der gesuchten arithmetischen Zahlenreihe, so ist $z_2 = z_1 + (m+1)\overline{d} = z_1 + d$, d.h., $\overline{d} = \frac{d}{(m+1)}$.

Entsprechend dem „Prinzip vom unzureichenden Grunde“ ist zunächst die Annahme, dass die Steuerpflichtigen im geschlossenen Intervall i (Klasse) mit den Intervallgrenzen $[a_i,b_i]$ gleichverteilt sind, Ausgangspunkt der Überlegungen.[73] In diesem Fall entspricht das durchschnittliche zu versteuernde Einkommen einer Klasse der Klassenmitte $\overline{zvE_i} = \frac{(a_i + b_i)}{2}$. Die Summe der zu versteuernden Einkommen aller Steuerpflichtigen einer Klasse ergibt sich entsprechend der Gleichung (1.2) aus dem Produkt des durchschnittlichen zu versteuernden Einkommens und der Anzahl der Steuerpflichtigen dieser Klasse: $zvE_i = h_i \overline{zvE_i}$. Die aggregierte Einkommensteuer der Klasse ist in diesem Fall mittels der Gleichungen (1.3) leicht zu ermitteln, da die absolute Häufigkeit der Steuerpflichtigen für jedes zu versteuernde Einkommen im Intervall gleich hoch ist $h_{(a_i)} = h_{(a_{i+1})} = \ldots = h_{(b_{i-1})} = h_{(b_i)} = \frac{h_i}{(b_i - a_i + 1)}$.

In der Regel entspricht das tatsächliche durchschnittliche zu versteuernde Einkommen einer Klasse jedoch nicht der Klassenmitte. Da in diesem Fall offensichtlich keine Gleichverteilung der Steuerpflichtigen vorliegt, muss eine Annahme über die Verteilung der Steuerpflichtigen innerhalb der Klasse getroffen werden.

Ausgehend von der Gleichverteilungsannahme wird in Abhängigkeit von der Lage des durchschnittlichen zu versteuernden Einkommens der Klasse zur Klassenmitte eine streng monoton steigende bzw. fallende diskrete Funktion (arithmetische Zahlenreihe) als Verteilung der Steuerpflichtigen in der Klasse unterstellt. Wird angenommen, dass die Anzahl der Steuerpflichtigen in der Klassenmitte dem Quotienten aus der Gesamtanzahl der Steuerpflichtigen dieser Klasse und der Klassenbreite entspricht, d.h. $h_{\left(\frac{a_i+b_i}{2}\right)} = \frac{h_i}{(b_i - a_i + 1)}$, reduziert sich das Problem darauf, eine bestimmte Anzahl von Steuerpflichtigen zwischen der unteren und oberen Klassenhälfte so umzuverteilen, dass die Summe der Einkommen der Klasse dem empirischen Wert entspricht. Diese Umverteilung wird zunächst derart normiert, dass die Anzahl der Steuerpflichtigen am Klassenanfang und Klassenende um genau zwei Steuerpflichtige differiert,

[73] Vgl. auch *Gyárfás/Quinke*, Allgemeines Statistisches Archiv 1993, S. 149 (150).

d.h. $\left|h_{(a_i)} - h_{(b_i)}\right| = 2$. Damit unterscheidet sich die Anzahl der Steuerpflichtigen in der Klassenmitte von der Anzahl der Steuerpflichtigen am Klassenanfang bzw. -ende um genau einen Steuerpflichtigen: $\left|h_{(a_i)} - h_{\left(\frac{a_i+b_i}{2}\right)}\right| = \left|h_{(b_i)} - h_{\left(\frac{a_i+b_i}{2}\right)}\right| = 1$.[74] Innerhalb der Klasse steigt bzw. sinkt mit jeder Änderung des zu versteuernden Einkommens zvE um eine Geldeinheit (DM bzw. €) die Anzahl der Steuerpflichtigen um $\frac{2}{(b_i - a_i)}$.[75] Mit Hilfe dieser Normierung wird die Anforderung der strengen Monotonie erfüllt. Die Ausprägung des Ausmaßes der Umverteilung innerhalb einer Klasse u_i kann mittels des empirischen zu versteuernden Einkommens der Klasse wie folgt bestimmt werden:

(1.4)
$$u_i = \frac{zvE_i - \frac{a_i + b_i}{2} h_i}{\sum_{zvE=a_i}^{b_i} \frac{zvE - a_i - \frac{b_i - a_i}{2}}{\frac{b_i - a_i}{2}} zvE}.$$

Die Anzahl der Steuerpflichtigen mit einem bestimmten zu versteuernden Einkommen ergibt sich unter den getroffenen Annahmen aus:

(1.5)
$$h_{(zvE)} = \frac{h_i}{b_i - a_i + 1} + \frac{zvE - a_i - \frac{b_i - a_i}{2}}{\frac{b_i - a_i}{2}} u_i.$$

Mit Hilfe von u_i und h_i kann für jedes zu versteuernde Einkommen die Anzahl der Steuerpflichtigen $h_{(zvE)}$ und somit $h_{(zvE)}zvE$ und $h_{(zvE)}t_{(zvE)}$ geschätzt werden. Durch Einsetzen der mittels Gleichung (1.5) ermittelten Häufigkeiten der Steuerpflichtigen in die Gleichungen (1.2) und (1.3) ergibt sich für jede Klasse i zum einen die Summe aller zu versteuernden Einkommen zvE_i, die genau dem empirischen Wert aus der Einkommensteuerstatistik entspricht, da $h_{(zvE)}$ mittels zvE_i bestimmt wurde. Zum anderen lässt sich die Gesamt-Einkommensteuer T_i dieser Klasse schätzen.

Durch diese Vorgehensweise bei der Berechnung der aggregierten Einkommensteuer einer Klasse wird gegenüber dem Gruppenmodell von *Boss*[76] die systematische Unterschätzung wesentlich verringert. Wird die aggregierte Steuer einer Klasse wie bei *Boss* nur durch die Multiplikation der Einkommensteuer auf das durchschnittliche zu versteuernde Einkommen der Klasse mit der Anzahl der Steuerpflichtigen der Klasse

[74] Das Vorzeichen der Differenz der Anzahl der Steuerpflichtigen an der Klassengrenze und der in der Klassenmitte bestimmt sich aus der Lage des durchschnittlichen zu versteuernden Einkommens der Klasse $\overline{zvE_i}$ zur Klassenmitte $\frac{a_i + b_i}{2}$. Bei $\overline{zvE_i} > \frac{a_i + b_i}{2}$ sind $h_{(b_i)} - h_{\left(\frac{a_i+b_i}{2}\right)} = 1$ und $h_{(a_i)} - h_{\left(\frac{a_i+b_i}{2}\right)} = -1$, wogegen sich bei $\overline{zvE_i} < \frac{(a_i + b_i)}{2}$ die Differenzen $h_{(a_i)} - h_{\left(\frac{a_i+b_i}{2}\right)} = 1$ und $h_{(b_i)} - h_{\left(\frac{a_i+b_i}{2}\right)} = -1$ ergeben.

[75] Diese vereinfachende Vorgehensweise führt dazu, dass im Modell die Anzahl der Steuerpflichtigen mit einem bestimmten Einkommen h_{zvE_i} nicht mehr in jedem Fall ganzzahlig ist.

[76] Vgl. Fn. 55.

bestimmt, erhält man bei einem progressiven Tarif lediglich die Untergrenze der möglichen Klassen-Gesamtsteuer. Verwendet man dagegen eine streng monotone diskrete Funktion, die auf Grundlage der empirisch ermittelten Anzahl der Steuerpflichtigen und der Summe des zu versteuernden Einkommens der Klasse bestimmt wird, so liegt die in der Simulationsrechnung ermittelte Gesamtsteuer einer Klasse zwischen der theoretischen Unter- und Obergrenze der möglichen Klassen-Gesamtsteuer.

Mit dieser Methode der Gruppensimulation lassen sich mit relativ geringem Aufwand recht genaue Ergebnisse, insbesondere bei der Tarifsimulation erzielen. Die Güte dieses Simulationsansatzes zeigt sich bei einer Gegenüberstellung der Ergebnisse der vom Statistischen Bundesamt durchgeführten Mikrosimulationen mit denen des hier vorgestellten diskreten Gruppensimulationsmodells[77].

Die zum Vergleich herangezogenen Simulationsrechnungen des Statistischen Bundesamtes wurden auf Basis von Einzeldatensätzen einer 10%-Stichprobe aus der Einkommensteuerstatistik 1995 durchgeführt. Bei der 10%–Stichprobe i.S. des § 7 Abs. 4 StStatG handelt es sich um eine formal anonymisierte Stichprobe aus der Gesamtheit der in der Einkommensteuerstatistik erfassten Einkommensteuerveranlagungen des Veranlagungszeitraumes 1995 die nach dem „Prinzip der vergleichbaren Präzision für gegliederte Ergebnisse“ gezogen wurde und als geschichtete Zufallsstichprobe im Statistischen Bundesamt vorliegt.[78]

Die Simulationsrechnungen erfolgten stark vereinfacht, d.h., es wurden nicht alle veranlagungsrelevanten Faktoren berücksichtigt. Auf das bei der Veranlagung für den Steuerpflichtigen ermittelte zu versteuernde Einkommen kam lediglich der Tarif gemäß § 32a EStG zur Anwendung, wobei die Berechnungen sowohl mit dem Steuertarif 1990 als auch mit dem Tarif 1996 durchgeführt wurden. Die für den Veranlagungszeitraum 1995 gültigen besonderen Tarifvorschriften, wie der Progressionsvorbehalt gemäß § 32b EStG, die Tarifbegrenzung bei gewerblichen Einkünften gemäß § 32c EStG, die Entlastung bei niedrigen Erwerbseinkommen gemäß § 32d EStG und die besonderen Steuersätze gemäß der §§ 34, 34b und 34c Abs. 4 EStG, blieben unberücksichtigt. Steuerermäßigungen und -erhöhungen sowie anrechenbare Steuern wurden ebenfalls nicht einbezogen. Die Stichprobenausgangswerte und -simulationsergebnisse wurden vom Statistischen Bundesamt auf die Grundgesamtheit hochgerechnet. Auf Basis der aggregierten Daten der hochgerechneten Stichprobenausgangswerte zur Anzahl der Steuerpflichtigen und zum aggregierten zu versteuernden Einkommen der Klassen sowie unter analoger Anwendung der Tarifvorschriften erfolgten auch die Berechnungen mittels des diskreten Gruppenmodells. Da bei der vom Statistischen Bundesamt vorgenommenen Klasseneinteilung die erste Klasse nach unten und die letzte Klasse nach oben offen sind, wurden diese Klassengrenzen für die Gruppensimulation heuristisch bestimmt. Dabei wurde – ausgehend von der Gleichverteilungsannahme – das durchschnittliche zu versteuernde Einkommen der Klasse mit der Klassenmitte gleichgesetzt $\overline{zvE_i} = \frac{a_i + b_i}{2}$. Die Intervallobergrenze entspricht dann dem Zweifachen der Klassenmitte, d.h. $\overline{zvE_i} \times 2 = b_i$. Dies gilt auch für die Ermittlung der Klassenuntergrenze

[77] Die Programmierung des diskreten Gruppensimulationsmodells erfolgte mit der Software „Mathematica“ Version 4.0.1.0. Die Berechnungen wurden auf einem Personalcomputer mit einem Pentium-III-Prozessor (733 MHz) durchgeführt.
[78] Zur Bildung der Stichprobe vgl. *Zwick*, WiSta 1998, S. 566 (570 – 573), und *Zwick*, (Fn. 24), S. 261 – 264.

der ersten Klasse, d.h. $-\overline{zvE_i} \times 2 = a_i$, da diese Klasse alle Steuerpflichtigen mit einem zu versteuernden Einkommen kleiner als eine DM beinhaltet und somit das zu versteuernde Einkommen in dieser Klasse negative Werte annimmt.

Die in den Tabellen 1 dargestellten Ergebnisse zeigen, dass die Differenzen zwischen den Ergebnissen der eigenen Berechnungen mittels des Gruppensimulationsmodells und denen der auf Basis der Einzeldatensätze erfolgten Simulationsrechnungen des Statistischen Bundesamtes (StBA) bei der Anwendung des Grundtarifs sowohl für die Einkommensteuertarife 1990 und 1996 sehr gering sind. Dies gilt auch für die Anwendung des Splittingtarifs.[79] Die Abweichungen liegen erwartungsgemäß weit unter den theoretischen Oberschranken für die relative Unterschätzung der Steuerschuld bei alleiniger Verwendung des Klassenmittelwerts.[80] Dabei ist bemerkenswert, dass sich diese Güte nicht nur beim Vergleich des Gesamtaufkommens, sondern auch für fast jede einzelne Klasse ergibt. Die bei einigen Modellen auftretenden Abweichungen in den unteren und oberen Einkommensklassen[81] sind bei der Tarifsimulation mittels des diskreten Gruppensimulationsmodells gering. Zudem muss erwähnt werden, dass die Güte der Ergebnisse des diskreten Gruppensimulationsmodells nicht wesentlich von den vom Statistischen Bundesamt gewählten Klassengrenzen – die in den vorliegenden Fällen unter Berücksichtigung der Rundungsregeln für die Bemessungsgrundlage[82] genau den Grundfreibeträgen der jeweiligen Tarife entsprechen – beeinflusst wird. Auch bei Simulationsrechnungen mit Tarifen, deren Grundfreibetrag nicht mit einer vom Statistischen Bundesamt festgesetzten Klassengrenze übereinstimmt, ergaben sich nur Differenzen in den dargestellten Größenordnungen.

[79] Für den Splittingtarif wurde auf die ausführliche Darstellung der Ergebnisse des Vergleichs verzichtet.
[80] Siehe *Gyárfás* (Fn. 43) , Tabelle 1 auf S. 34.
[81] Siehe z.B. *Spahn* (Fn. 34), S. 129 und 133, sowie *Petersen* (Fn. 64), S. 150.
[82] Für den Tarif 1990 und 1996 ist gemäß § 32a Abs. 2 EStG das zu versteuernde Einkommen auf den nächsten durch 54 ohne Rest teilbaren vollen DM-Betrag abzurunden, wenn es nicht bereits durch 54 ohne Rest teilbar ist.

Tabelle1 : Vergleich der Berechnung des Steueraufkommens mittels Mikrosimulation und Gruppensimulation – Einkommensverteilung 1995 / Grundtarif

lfd. Nr.	zu versteuerndes Einkommen *von ... bis unter ... DM*		Tarif 90			Tarif 96
			berechnete ESt in TDM			
			Simulation StBA	*eigene Berechnungen*	*relative Abweichung*	*relative Abweichung*
1		unter 1	-	-	0,0000%	0,0000%
2	1 -	5.670	-	-	0,0000%	0,0000%
3	5.670 -	8.154	180.739	180.734	-0,0028%	0,0000%
4	8.154 -	12.096	965.929	965.970	0,0042%	0,0000%
5	12.096 -	12.366	83.707	83.706	-0,0012%	-0,0685%
6	12.366 -	13.068	219.226	219.228	0,0009%	0,0083%
7	13.068 -	18.036	1.837.818	1.837.820	0,0001%	0,0022%
8	18.036 -	25.002	4.018.826	4.018.823	-0,0001%	0,0008%
9	25.002 -	30.023	4.408.774	4.408.747	-0,0006%	-0,0003%
10	30.023 -	40.013	14.380.662	14.381.434	0,0054%	0,0034%
11	40.013 -	50.004	15.660.133	15.660.329	0,0013%	0,0010%
12	50.004 -	55.728	7.173.713	7.173.689	-0,0003%	-0,0003%
13	55.728 -	58.644	3.108.599	3.108.590	-0,0003%	-0,0003%
14	58.644 -	60.048	1.403.312	1.403.312	0,0000%	0,0001%
15	60.048 -	66.366	5.465.481	5.465.480	0,0000%	-0,0005%
16	66.366 -	70.038	2.652.440	2.652.457	0,0006%	0,0006%
17	70.038 -	75.006	3.013.301	3.013.302	0,0000%	0,0001%
18	75.006 -	100.008	8.784.647	8.782.625	-0,0230%	-0,0230%
19	100.008 -	120.042	3.299.181	3.298.992	-0,0057%	-0,0058%
20	120.042 -	240.084	6.458.174	6.458.172	0,0000%	0,0000%
21	240.084 -	480.168	3.488.455	3.488.455	0,0000%	0,0003%
22	480.168 -	1.000.026	2.549.395	2.549.390	-0,0002%	0,0000%
23	1.000.026 -	und mehr	7.226.559	7.226.536	-0,0003%	-0,0003%
24		**gesamt**	**96.379.068**	**96.377.792**	**-0,0013%**	**-0,0018%**

Quelle: Statistisches Bundesamt, Wiesbaden; eigene Berechnungen.

Die bei der Tarifsimulation erzielte Genauigkeit müsste sich auch bei der Berechnung der Aufkommenswirkungen von fixen Beträgen, die bei der Ermittlung der Bemessungsgrundlage abgezogen werden (z.B. Pauschbeträgen),[83] mittels des diskreten Gruppensimulationsmodells erreichen lassen.[84] Bei derartigen Simulationsrechnungen können sich jedoch die Differenzen zu den Ergebnissen einer Mikrosimulation vergrößern, wenn der untersuchte steuerliche Fixbetrag nicht bei allen Steuerpflichtigen zum Abzug kommt und die Verteilung der Steuerpflichtigen, bei denen der Abzug erfolgt, nicht der Verteilung aller Steuerpflichtigen entspricht. Um in diesem Fall die Güte der Ergebnisse zu verbessern, müssten Informationen über die Verteilung der Steuerpflichtigen, bei denen der zu untersuchende steuerliche Fixbetrag abzuziehen ist – soweit sie zur Verfügung stehen –, bei der Simulationsrechnung berücksichtigt werden. Aus der veröffentlichten Einkommensteuerstatistik lassen sich – wie bereits ausgeführt – für die meisten steuerlichen Abzüge für jede Klasse die Anzahl der Steuerpflichtigen und der Gesamtbetrag in TDM entnehmen. Somit liegen zwar Informationen über die Verteilung auf die einzelnen Einkommensklassen,

[83] Im Folgenden wird ausschließlich die Problematik der Abzüge von der steuerlichen Bemessungsgrundlage behandelt, deren Aufkommenswirkungen durch eine entsprechende Bemessungsgrundlagenerhöhung in der Simulationsrechnung bestimmt werden. Die Ausführungen gelten aber prinzipiell auch für Regelungen, die zu einer Erhöhung der steuerlichen Bemessungsgrundlage geführt haben und deren Aufkommenswirkungen durch eine simulierte Bemessungsgrundlagenverminderung ermittelt wird.

[84] Da für einen derartigen Sachverhalt keine Mikrosimulation vom Statistischen Bundesamt durchgeführt wurde, kann kein Vergleich mit der Gruppensimulation anhand eines konkreten Beispiels dargestellt werden.

nicht aber über die Verteilung der Steuerpflichtigen innerhalb der Klassen vor. Enthalten die Tabellen der Einkommensteuerstatistik keine Angaben über das zu versteuernde Einkommen der Steuerpflichtigen, bei denen der entsprechende Betrag zum Abzug kam, so muss für die Gruppensimulationsrechnung auf die Summe der zu versteuernden Einkommen aller Steuerpflichtigen und damit auf die bei der Gruppensimulation daraus abgeleitete Verteilung von allen Steuerpflichtigen in dieser Klasse zurückgegriffen werden. Dies kann allerdings zu einer größeren Abweichung von den Ergebnissen einer Mikrosimulationsrechnung führen.

Unter Verwendung der unter IV.1. definierten Symbole lässt sich diese Problematik formal wie folgt darstellen: Aus den aggregierten Daten der Einkommensteuerstatistik sind für jede Klasse i die Häufigkeit G_l des Vorliegens eines steuerlichen Sachverhalts (Anzahl der Steuerpflichtigen, die diesen Tatbestand erfüllen) und die Höhe seiner Ausprägung G_l in TDM bekannt, wobei der durchschnittliche Wert einer Klasse $\overline{G_i} = \frac{G_i}{g_i}$ ist. Im Fall eines fixen steuerlichen Abzugsbetrags ist G_l für jede Klasse gleich hoch. Die finanziellen Auswirkungen der steuerlichen Regelung je Klasse ergeben sich aus der Differenz ?T_i der jeweiligen Summe der Einkommensteuer der Klasse mit (T_i^g) und ohne (T_i) Berücksichtigung des Abzugs von der steuerlichen Bemessungsgrundlage:

$\Delta T_i = T_i^g - T_i$,

wobei

$T_i^g = h_{(zvE)}^g t_{(zvE+\overline{G_i})}$.

Dabei sind $h_{(zvE)}^g$ die Anzahl der Steuerpflichtigen mit einem bestimmten zvE, die in G_l enthalten sind, und $t_{(zvE+\overline{G_i})}$ die Einkommensteuer auf die um $\overline{G_i}$ erhöhte Bemessungsgrundlage zvE. Die Genauigkeit der Simulationsrechnung wird – wie bereits ausgeführt – auch davon beeinflusst, ob für die Klasse i die Summe der zu versteuernden Einkommen der Steuerpflichtigen, bei denen der untersuchte steuerliche Betrag (zvE_i^g) zum Abzug kommt, bekannt ist. Für ui und $h_{(zvE)}^g$, die mittels der Gleichungen (1.4) und (1.5) bestimmt werden, ist es relevant, ob sie auf Grundlage des zu versteuernden Einkommens aller Steuerpflichtigen (zvE_i) oder mittels des zu versteuernden Einkommens der in G_l enthaltenen Steuerpflichtigen (zvE_i^g) ermittelt werden. Ist zvE_i^g Bekannt, entspricht $h_i^g = g_i$, während im anderen Fall ui auf Grundlage von zvE_l ermittelt wird und vereinfachend davon ausgegangen werden muss, dass $h_{(zvE)}^g = h_{(zvE)} \frac{g_i}{h_i}$ beträgt. Bei dieser Vorgehensweise wird bei allen untersuchten Sachverhalten eine für die jeweilige Klasse identische Verteilung der Steuerpflichtigen mit einem bestimmten zu versteuernden Einkommen unterstellt.

Eine weitere Einschränkung der Genauigkeit ergibt sich bei der Anwendung des diskreten Gruppensimulationsmodells zur Bestimmung der Aufkommenswirkungen von Regelungen, die Abzüge von der Bemessungsgrundlage betreffen, die bei jedem einzelnen Steuerpflichtigen unterschiedlich hoch sein

können – wie dies z.B. bei Abschreibungen oder der Verlustverrechnung der Fall ist. Auch wenn sich die Verteilung der Steuerpflichtigen und des Abzugsbetrags auf die einzelnen Einkommensklassen sowie die Summe der zu versteuernden Einkommen der betreffenden Steuerpflichtigen aus den Tabellen der Einkommensteuerstatistik entnehmen lassen, muss bei einer Gruppensimulation für die Verteilung des untersuchten variablen Abzugsbetrags innerhalb der jeweiligen Einkommensklasse – analog zur Verteilung des zu versteuernden Einkommens – eine Annahme getroffen werden, da sich die tatsächliche Verteilung auch in diesem Fall nicht aus den aggregierten Daten ermitteln lässt. Für die hier durchgeführten Berechnungen wird wieder vereinfachend von einer Gleichverteilung ausgegangen, so dass bei jedem Steuerpflichtigen einer Klasse der aus der Summe des zu untersuchenden Abzugsbetrags der jeweiligen Klasse bestimmte Durchschnittswert $\overline{G_i} = \frac{G_i}{g_i}$ zum Ansatz kommt. Im Unterschied zu einem fixen Abzugsbetrag kann sich jedoch für jede Klasse ein unterschiedlich hoher Durchschnitt $\overline{G_i}$ ergeben.

In Tabelle 2 werden die Ergebnisse der Mikrosimulationsrechnungen des Statistischen Bundesamtes zur Berechnung des Aufkommens der Einkommensteuer für einen eingeschränkten Verlustausgleich denen des diskreten Gruppenmodells gegenübergestellt.[85] Die Einschränkung des Verlustausgleichs bestand in der Versagung des vertikalen Verlustausgleichs mit positiven Einkünften aus anderen Einkunftsarten.

Bei der Berechnung der Einkommensteuer kam wie bereits bei den in den Tabellen 1 dargestellten Vergleichen lediglich die Tarifvorschrift des § 32a EStG zur Anwendung. Jedoch wurden diese Simulationsrechnungen nur mit dem Tarif 1990 durchgeführt. In Tabelle 2 wird die relative Abweichung der mittels Gruppensimulation berechneten Einkommensteuer von der auf Grundlage der Mikrosimulation berechneten Einkommensteuer – getrennt nach Grund- und Splittingtarif – sowohl für die einzelnen Einkommensklassen als auch insgesamt dargestellt. Zudem wird am Ende der Tabelle die relative Abweichung zwischen den mittels der verschiedenen Simulationsmethoden berechneten finanziellen Auswirkungen einer Versagung des vertikalen Verlustausgleichs angegeben. Die finanziellen Auswirkungen ergeben sich aus der Differenz der jeweiligen berechneten Summen der Einkommensteuer aller Steuerpflichtigen mit negativen Einkünften für den Fall des unbeschränkten und des eingeschränkten Verlustausgleichs.

Die Gruppensimulation wurde zudem auf Grundlage unterschiedlich aggregierter Daten durchgeführt. In einem Fall wurden tabellierte Daten aus der vom Statistischen Bundesamt hochgerechneten Stichprobe verwendet, die ausschließlich Angaben zu den Steuerpflichtigen mit negativen Einkünften enthielten, d.h., die Summe des zu versteuernden Einkommens dieser Steuerpflichtigen je Klasse (zvE_i^g) war bekannt. Dagegen wurde im anderen Fall bei der Gruppensimulationsrechnung aus den tabellierten Stichprobendaten nur die Summe des zu versteuernden Einkommens aller Steuerpflichtigen der Klasse (zvE_i) zur Simulation der Verteilung der Bemessungsgrundlage innerhalb der Klasse verwendet.

[85] Die Nummerierung der Klassen in Tabelle 3 entspricht der in den Tabellen 1 und 2.

Tabelle 2: Vergleich der Berechnung des Steueraufkommens und der finanziellen Auswirkungen bei Versagung des vertikalen Verlustausgleichs mittels Mikro- und Gruppensimulation auf Grundlage von zvE-Tabellen – Einkommensverteilung 1995

zvE-Klasse Nr.	relative Abweichung zwischen den Simulationsergebnissen des Statistischen Bundesamts und denen des diskreten Gruppenmodells			
	Interpolation mittels zvE_i^g		Interpolation mittels zvE_i	
	Grundtarif	Splittingtarif	Grundtarif	Splittingtarif
1	-1,4182%	-100,0000%	194,1274%	263,1967%
2	-34,5701%	-35,6779%	-37,4919%	-38,2353%
3	-20,3230%	-17,8262%	-19,9858%	-17,7484%
4	-14,7191%	-11,8263%	-15,0016%	-11,6785%
5	-9,4212%	-7,9040%	-9,4722%	-7,8956%
6	-11,7048%	-7,3189%	-11,7094%	-7,3079%
7	-9,8775%	-6,5023%	-10,5133%	-6,4304%
8	-6,3505%	-4,3696%	-6,3986%	-4,9910%
9	-5,1233%	-2,9831%	-5,2855%	-3,0239%
10	-3,6476%	-2,7450%	-3,6717%	-3,1454%
11	-3,3690%	-2,6741%	-3,8018%	-3,1458%
12	-2,4027%	-2,4326%	-2,7062%	-2,4803%
13	-2,7901%	-2,3300%	-2,8571%	-2,3755%
14	-2,3323%	-2,4959%	-2,4071%	-2,5000%
15	-2,6210%	-2,4570%	-2,7292%	-2,5696%
16	-2,5111%	-2,2322%	-2,7662%	-2,2305%
17	-2,0388%	-2,1054%	-2,1141%	-2,1570%
18	-1,6480%	-1,7187%	-2,6890%	-2,5163%
19	-0,1687%	-0,1625%	-0,5135%	-0,3792%
20	-0,1687%	-0,0181%	-2,3527%	-0,7592%
21	-0,0071%	-0,0080%	-0,5272%	-0,1134%
22	-0,0038%	-0,0037%	0,0445%	0,2148%
23	-0,0010%	-0,0010%	-5,2218%	0,0058%
gesamt	-1,9124%	-3,7316%	2,6085%	3,0275%
gesamt ohne Klasse 1	1,9282%	-1,8186%	-3,5040%	-2,1424%
finanzielle Auswirkung	-6,8567%	-14,5591%	14,6799%	13,0822%
finanzielle Auswirkung ohne Klasse 1	-7,5351%	-7,4848%	-7,7054%	-7,6267%

Quelle: Statistisches Bundesamt, Wiesbaden; eigene Berechnungen.

Ein Vergleich der Ergebnisse in den Tabellen 1 und 2 zeigt, dass die Abweichungen der Ergebnisse der Gruppensimulation von denen der Mikrosimulation bei einer Simulation der Aufkommenswirkungen von Abzügen von der Bemessungsgrundlage, die bei jedem Steuerpflichtigen individuell verschieden sein können, gegenüber einer Tarifsimulation wesentlich höher ausfallen. Für die mittels der Gruppensimulation berechnete aggregierte Einkommensteuer aller Steuerpflichtigen mit negativen Einkünften ergeben sich bei einer Interpolation der Verteilung mittels der Klassensumme des zu versteuernden Einkommens der Steuerpflichtigen mit negativen Einkünften (zvE_i^g) relative Abweichungen gegenüber der Mikrosimulation

von -1,9% (Grundtarif) und -3,7% (Splittingtarif) und 2,6% (Grundtarif) sowie 3,0% (Splittingtarif) bei der Verwendung der Summe des zu versteuernden Einkommens aller Steuerpflichtigen der Klasse (zvE_i).

Die Differenzen sind in den unteren Einkommensklassen am größten und nehmen mit zunehmender Höhe der Bemessungsgrundlage tendenziell ab. Die größten relativen Abweichungen ergeben sich für die Klasse 1, in der die Steuerpflichtigen ein zu versteuerndes Einkommen kleiner als eine DM aufweisen. Da diese Klasse in der Einkommensteuerstatistik nicht weiter untergliedert wird, ist das Gruppensimulationsmodell in diesem Bereich sehr ungenau. Dies hat zur Folge, dass die Schätzung der Anzahl der Steuerpflichtigen, die nach der Simulation auf Grund der Versagung des vertikalen Verlustausgleichs ein positives Einkommen oberhalb des Grundfreibetrages aufweisen, sehr unsicher ist. Das Ergebnis in dieser Klasse ist zudem von der Höhe der Klassenuntergrenze abhängig, die – wie bereits ausgeführt – heuristisch bestimmt werden muss. Die Einbeziehung der Klasse der Steuerpflichtigen, die ein zu versteuerndes Einkommen kleiner als eine DM aufweisen, erscheint bei einer Untersuchung der Aufkommenswirkungen einer simulierten Bemessungsgrundlagenerhöhung – soweit insbesondere diese Steuerpflichtigen davon betroffen sind – nur bei einer Mikrosimulation sinnvoll. Auf Grund der mangelnden Daten kann ein Gruppenmodell in diesem Fall sich nur zufällig dem Ergebnis der Mikrosimulation nähern. Wird die Klasse der Steuerpflichtigen, die ein zu versteuerndes Einkommen kleiner als eine DM aufweisen, bei dem Vergleich der Simulationsverfahren nicht berücksichtigt, betragen die relativen Abweichungen zwischen den berechneten Steueraufkommen für alle Steuerpflichtigen mit negativen Einkünften bei der Interpolation mittels zvE_i^g 1,9% (Grundtarif) und –1,8% (Splittingtarif) sowie bei der Verwendung von zvE_i -3,5% (Grundtarif) und -2,1% (Splittingtarif).

Da sich das auf die unmodifizierte Bemessungsgrundlage (Tabellen 1) mittels Mikrosimulation berechnete Steueraufkommen nicht so stark von dem der Gruppensimulation unterscheidet wie die Steueraufkommen bei einer modifizierten Bemessungsgrundlage (Tabelle 2), weisen die finanziellen Auswirkungen der Bemessungsgrundlagenmodifikation wesentlich höhere relative Abweichungen zwischen der Mikrosimulation und der Gruppensimulation auf. Dabei fallen die Differenzen in den unteren Einkommensklassen besonders ins Gewicht. Die relativen Abweichungen zwischen der Mikrosimulation und der Gruppensimulation betragen für die gesamten finanziellen Auswirkungen bei der Einbeziehung aller Klassen -6,9% (Grundtarif) und -14,6% (Splittingtarif) bei Anwendung von zvE_i^g sowie 14,7% (Grundtarif) und 13,1% (Splittingtarif) unter Zugrundelegung von zvE_i. Ohne Berücksichtigung der untersten Einkommensklasse ergeben sich relative Abweichungen von etwa –7,5% (Grund- und Splittingtarif) bei der Interpolation mittels zvE_i^g und -7,7% (Grundtarif) bzw. -7,6% (Splittingtarif) bei der Verwendung von zvE_i. Dass die Gruppensimulation ohne Einbeziehung der ungenauen Werte der ersten Klasse grundsätzlich zu einer Unterschätzung der finanziellen Auswirkungen führt, entspricht den Erwartungen, da unter Zugrundelegung des durchschnittlichen Abzugsbetrags bei jedem betroffenen Steuerpflichtigen – wie schon unter Verwendung von Durchschnittswerten des zu versteuernden Einkommens im Lohnsteuermodell von *Boss*[86] – die Untergrenze des möglichen Steuerminderaufkommens bestimmt wird.

[86] Vgl. Fn. 55.

Weiterhin zeigt ein Vergleich der Ergebnisse der Tabelle 2, dass die Gruppensimulationsergebnisse, die auf Grundlage der Klassensumme des zu versteuernden Einkommens der Steuerpflichtigen mit negativen Einkünften (zvE_i^g) berechnet wurden, wie erwartet tendenziell geringere Abweichungen zu den Mikrosimulationsergebnissen aufweisen als die auf Grundlage der Klassensumme des zu versteuernden Einkommens aller Steuerpflichtigen der Klasse (zvE_i) ermittelten Gruppensimulationsergebnisse. Daraus kann aber nicht gefolgert werden, dass die Abweichungen immer in den hier ermittelten Größenordnungen auftreten, da sich die (unbekannte) Verteilung der Steuerpflichtigen innerhalb einer Klasse grundsätzlich in jeder Klasse und bei jedem untersuchten steuerlichen Sachverhalt anders darstellen kann. Dies wird auch bei einem Vergleich der einzelnen in Tabelle 2 dargestellten Klassenergebnisse deutlich.

IV.3. Daten nach Größenklassen des Gesamtbetrags des Einkommens

Der überwiegende Teil der Tabellen der Einkommensteuerstatistik, insbesondere der Nachweis des Umfangs der Anwendung vieler steuerlicher Einzelregelungen, ist nicht nach Größenklassen des zu versteuernden Einkommens, sondern nach Größenklassen des Gesamtbetrags der Einkünfte[87] gegliedert.[88]

In diesen Fällen kann die Datenbasis – wiederum jeweils für die Steuerpflichtigen, die mit dem Grundtarif, und die Steuerpflichtigen, die mit dem Splittingtarif besteuert werden – formal wie folgt dargestellt werden:

Gegeben ist eine Klasseneinteilung nach dem Gesamtbetrag der Einkünfte für j=1 bis m Klassen mit den Klassengrenzen [c_j, D_j], wobei c_1=-8 , D1=-1, c2=0 und dm=8 .[89]

Für jede Klasse j ist u.a. folgendes bekannt:

- die Klassenhäufigkeit H_j (Anzahl der Steuerpflichtigen der Klasse, für die ein zu versteuerndes Einkommen festgestellt wurde),
- die Häufigkeit g_j des Vorliegens eines steuerlichen Sachverhalts (Anzahl der Steuerpflichtigen, die diesen Tatbestand erfüllen) und die Höhe seiner Ausprägung G_j (in TDM bzw. T€),
- die Summe der zu versteuernden Einkommen aller Steuerpflichtigen dieser Klasse zvE_j und
- die Summe der festgesetzten Einkommensteuer aller Steuerpflichtigen dieser Klasse ESt_j.

Bei der Anwendung des Gruppensimulationsmodells auf der Basis von Daten, die nur in der Form von nach dem Gesamtbetrag der Einkünfte (GdE) gegliederten Tabellen zur Verfügung stehen, ergibt sich das Problem, dass die Verteilung der Steuerpflichtigen mit einem bestimmten zu versteuernden Einkommen ($h_{(zvE)}$) nur schwer geschätzt werden kann. Dies liegt darin begründet, dass für die Steuerpflichtigen einer

[87] Der Gesamtbetrag der Einkünfte ergibt sich gemäß § 2 Abs. 3 Satz 1 EStG aus der Summe der Einkünfte, vermindert um den Altersentlastungsbetrag (§ 24a EStG) und den Freibetrag für Land- und Forstwirte (§ 13 Abs. 3 EStG). Vgl. dazu auch R 3 EStR.

[88] Siehe dazu auch Abschnitt II.

[89] Für die Klassen j>1 gilt zudem, dass die Steuerpflichtigen ein zu versteuerndes Einkommen größer als null DM haben. Die erste Klasse (*j=1*) enthält die so genannten Verlustfälle, die vorliegen, wenn der Steuerpflichtige mit einem negativen Einkommen veranlagt wurde. Dabei kann sich bereits bei der Ermittlung der Summe der Einkünfte oder erst bei der Ermittlung des zu versteuernden Einkommens – z.B. durch den Abzug von Sonderausgaben und außergewöhnlichen Belastungen – ein negativer Wert ergeben haben (vgl. *Statistisches Bundesamt*, Finanzen und Steuern, Reihe 7.1 Lohn- und Einkommensteuer 1995, S. 10). Dass heißt, die Klasse der Verlustfälle kann Steuerpflichtige enthalten, die einen positiven Gesamtbetrag der Einkünfte aufweisen, aber aufgrund ihres negativen Einkommens nicht in den Klassen für positive Gesamtbeträge der Einkommen (*j>1*) enthalten sind.

GdE-Klasse j lediglich das durchschnittliche zu versteuernde Einkommen der Klasse $\overline{zvE_j} = \frac{zvE_j}{h_j}$ unmittelbar zu bestimmen ist; die Intervallbreite des möglichen zu versteuernden Einkommens $[a_i, b_i]$ kann für diese Steuerpflichtigen dagegen nicht aus den Tabellenwerten abgeleitet werden.

Aus der Zuordnung eines Steuerpflichtigen zu einer bestimmten GdE-Klasse könnte lediglich die zvE-Obergrenze b_i in Form des theoretisch höchstmöglichen zu versteuernden Einkommens der Klasse bestimmt werden, indem die GdE-Klassenobergrenze d_j um die steuerlichen Mindestabzüge wie z.B. den Sonderausgabenpauschbetrag nach § 10c Abs. 1 EStG vermindert wird. Eine theoretische Untergrenze für das zu versteuernde Einkommen (a_j) lässt sich dagegen nicht ermitteln, da das zu versteuernde Einkommen durch diverse Abzüge vom Gesamtbetrag der Einkünfte, z.B. Sonderausgaben, Verlustabzug nach § 10d EStG oder Aufwendungen für außergewöhnliche Belastungen, jeden beliebigen Wert unterhalb der GdE-Klassenobergrenze d_j annehmen kann. Deshalb müsste in diesem Fall der Intervallanfang a_i (kleinstmögliches zu versteuerndes Einkommen) großzügig geschätzt werden, was eine höhere Ungenauigkeit der Simulationsergebnisse mit sich brächte. Um die aus diesem Informationsdefizit resultierenden Abweichungen bei den mittels des Gruppenmodells durchzuführenden Simulationsrechnungen zu verringern, wurden vom Statistischen Bundesamt für die im Rahmen dieser Arbeit durchzuführenden Simulationsrechnungen Kreuztabellen erstellt, mit deren Hilfe ein Teil der aggregierten Daten der Einkommensteuerstatistik, die nach dem „Gesamtbetrag der Einkünfte“ gruppiert sind, in eine Darstellung, gegliedert nach dem Größenklassenmerkmal „zu versteuerndes Einkommen“, überführt werden kann.

In diesen Kreuztabellen werden die absolute Häufigkeit der Steuerpflichtigen h_i mit einem zu versteuernden Einkommen in der Klasse *i* sowie deren Summe des zu versteuernden Einkommens zvE_i mit der absoluten Häufigkeit der Steuerpflichtigen h_j mit einem Gesamtbetrag der Einkünfte in der Klasse *j* und der Summe der zu versteuernden Einkommen zvE_j dieser Steuerpflichtigen zusammengeführt. Dadurch ergibt sich eine Matrix der absoluten Häufigkeiten der Steuerpflichtigen h_{ij} und der dazugehörigen Summen der zu versteuernden Einkommen zvE_{ij}.

Mit Hilfe dieser Matrix lässt sich die Verteilung der Steuerpflichtigen mit einem bestimmten zu versteuernden Einkommen aus den aggregierten Daten der Einkommensteuerstatistik, die nach dem Größenklassenmerkmal „Gesamtbetrag der Einkünfte“ gruppiert sind, analog zu den der Gleichungen (1.4) und (1.5) wie folgt schätzen:

$$h_{(zvE)} = \frac{h_{ij}}{(b_i - a_i + 1)} + \frac{zvE - a_i - \left(\frac{b_i - a_i}{2}\right)}{\left(\frac{b_i - a_i}{2}\right)} u_{ij}, \tag{1.8}$$

wobei

(1.9)
$$u_{ij} = \frac{zvE_{ij} - \frac{a_i + b_i}{2} h_{ij}}{\sum_{zvE=a_i}^{b_i} \frac{zvE - a_i - \frac{b_i - a_i}{2}}{\frac{b_i - a_i}{2}}} zvE \,.$$

Unter Verwendung des diskreten Gruppensimulationsmodells kann das Aufkommen der Einkommensteuer dann mittels der Gleichung (1.3) bestimmt werden.

Ein Vergleich der Ergebnisse der auf den GdE-Tabellen der Einkommensteuerstatistik basierenden Gruppensimulation mit denen der Mikrosimulation zeigt, dass unter Verwendung der vom Statistischen Bundesamt erstellten Kreuztabellen eine Tarifsimulation mit dem diskreten Gruppenmodell wiederum recht genaue Ergebnisse liefert. Die relativen Abweichungen zwischen den Gruppen- und Mikrosimulationsergebnissen sind in Tabelle 3 enthalten. Die Simulationsrechnungen wurden mit dem Einkommensteuertarif 1990 durchgeführt und erfolgten wiederum ausschließlich unter Anwendung der Tarifvorschrift des § 32a EStG. Bei den Steuerpflichtigen, die nach dem Grundtarif besteuert werden, ergibt sich mit einer Abweichung von 0,027% eine geringfügige Überschätzung der Einkommensteuer bei der Gruppensimulation. Beim Splittingtarif liegen die Gruppensimulationsergebnisse mit -0,048% leicht unter denen der Mikrosimulation.

Die bei der Gruppensimulation aufgetretene geringfügige Überschätzung beim Grundtarif ist auf die verwendeten Daten zur steuerlichen Bemessungsgrundlage zurückzuführen. In der Tabelle 3 ist dem Vergleich der ermittelten aggregierten Einkommensteuer ein Vergleich der für die jeweiligen Simulationsrechnungen zu Grunde gelegten Daten zur steuerlichen Bemessungsgrundlage vorangestellt. Dieser Vergleich zeigt, in welchem Umfang die aus der Stichprobe hochgerechneten Ausgangsdaten für die Mikrosimulation von den bei der Gruppensimulation zu Grunde gelegten Daten der Einkommensteuerstatistik abweichen. Es wird deutlich, dass die Differenzen zwischen der vom Statistischen Bundesamt hochgerechneten Stichprobe und den Werten der Grundgesamtheit der Einkommensteuerstatistik sehr gering sind und folglich die Simulationsergebnisse nur unwesentlich davon beeinflusst werden. In der Gruppensimulation auf Basis der GdE-Tabellen werden die fixen und variablen Abzüge von der Bemessungsgrundlage entsprechend der bisherigen Vorgehensweise berücksichtigt. Das Steueraufkommen, welches sich ohne die Berücksichtigung eines bestimmten steuerlichen Sachverhaltes (g, G) ergeben würde, kann analog zur Gleichung (1.7) geschätzt werden. Dabei muss i.d.R. von $h^g_{(zvE)} = h_{(zvE)} \frac{g_j}{h_j}$ ausgegangen werden, da das zu versteuernde Einkommen der Steuerpflichtigen, bei denen der Abzug von der Bemessungsgrundlage zur Anwendung kommt (zvE^g_j), nicht bekannt ist. Dies bedeutet wiederum, dass bei allen untersuchten Sachverhalten eine für die jeweilige GdE-Klasse identische Verteilung der Steuerpflichtigen mit einem bestimmten zu versteuernden Einkommen unterstellt wird. Diese vereinfachende Vorgehensweise kann bei der Verwendung von nach dem Gesamtbetrag der Einkünfte klassierten Tabellen als Datenbasis für die Gruppensimulationsrechnung im Vergleich zu Daten aus Tabellen, die nach dem zu versteuernden Ein-

kommen gegliedert sind, zu größeren Abweichungen führen, da die Unterstellung einer identischen Verteilung der Steuerpflichtigen mit einem bestimmten zu versteuernden Einkommen für ein Intervall des zu versteuernden Einkommens [a_i, b_i] prinzipiell zu geringeren Differenzen führt als für ein Intervall des Gesamtbetrags des Einkommens [c_j, d_j].

Tabelle 3: Vergleich der Ausgangsdaten der Stichprobe mit denen der Einkommensteuerstatistik und Vergleich der Ergebnisse der Tarifsimulation mittels Gruppenmodell auf Basis von GdE-Tabellen der Einkommensteuerstatistik mit denen der Mikrosimulation auf Basis einer Stichprobe – Einkommensverteilung 1995

	Abweichung zwischen					
	Stichprobe und Einkommensteuerstatistik				Mikrosimulations- und Gruppensimulationsergebnissen	
	Anzahl der Steuerpflichtigen		zu versteuerndes Einkommen		berechnete Einkommensteuer	
zvE-Klasse Nr.	Grd.-Tarif	Sp.-Tarif	Grd.-Tarif	Sp.-Tarif	Grd.-Tarif	Sp.-Tarif
1	1,4916%	-0,0093%	0,0934%	0,0690%	0,0000%	0,0000%
2	0,8065%	-0,0003%	0,4750%	0,0010%	0,0000%	0,0000%
3	0,0927%	-0,0020%	0,0891%	-0,0023%	0,0702%	-1,0534%
4	0,0692%	-0,0005%	0,0687%	-0,0009%	0,0723%	-0,2947%
5	-0,0458%	0,0226%	-0,0451%	0,0233%	-0,0455%	-0,1903%
6	0,0984%	-0,0266%	0,0993%	-0,0269%	0,1007%	-0,2267%
7	0,0482%	0,0040%	0,0493%	0,0036%	0,0502%	-0,1414%
8	0,0784%	-0,0015%	0,0780%	-0,0016%	0,0777%	-0,0970%
9	0,0649%	0,0030%	0,0659%	0,0030%	0,0657%	-0,0680%
10	0,0386%	-0,0010%	0,0391%	-0,0008%	0,0447%	-0,0544%
11	0,0324%	0,0001%	0,0325%	-0,0002%	0,0338%	-0,0445%
12	0,0332%	-0,0019%	0,0333%	-0,0023%	0,0329%	-0,0392%
13	0,0073%	0,0176%	0,0073%	0,0171%	0,0070%	-0,0172%
14	0,0788%	0,0000%	0,0790%	0,0005%	0,0792%	-0,0323%
15	0,0212%	0,0000%	0,0216%	0,0001%	0,0217%	-0,0311%
16	-0,0373%	0,0077%	-0,0384%	0,0084%	-0,0383%	-0,0203%
17	0,0761%	-0,0106%	0,0754%	-0,0105%	0,0751%	-0,0378%
18	0,0235%	0,0019%	0,0228%	0,0021%	-0,0006%	-0,0384%
19	0,0064%	-0,0073%	0,0057%	-0,0075%	-0,0004%	-0,0302%
20	-0,0046%	-0,0094%	-0,0130%	-0,0072%	-0,0162%	-0,0180%
21	0,0467%	0,0000%	0,0369%	0,0000%	0,0357%	-0,0048%
22	0,0000%	0,0000%	0,0000%	0,0000%	0,0000%	-0,0021%
23	0,0000%	0,0000%	0,0000%	0,0000%	0,0001%	-0,0003%
gesamt	**0,2782%**	**-0,0007%**	**0,0388%**	**-0,0020%**	**0,0269%**	**-0,0477%**

Quelle: Statistisches Bundesamt, Wiesbaden; eigene Berechnungen.

Dies bestätigt sich auch bei dem in Tabelle 4 dargestellten Vergleich der Ausgangsdaten und der Ergebnisse der Mikrosimulation mit denen einer Gruppensimulation auf Basis von nach dem Gesamtbetrag der Einkünfte gegliederten Tabellen für die Berechnung des Steueraufkommens bei der Versagung des vertikalen Verlustausgleichs zwischen den Einkunftsarten. Bei einer auf Basis von GdE-Tabellen durchgeführten Gruppensimulation weichen wegen der unterschiedlichen Datenbasen bereits die Bemessungsgrundlagen vor der Verlustsimulationsrechnung von denen der Mikrosimulation ab.

Dies ist darin begründet, dass das zu versteuernde Einkommen der Steuerpflichtigen mit negativen Einkünften (zvE_j^g) den GdE-Tabellen der Einkommensteuerstatistik nicht entnommen werden kann. Die Bemessungsgrundlage der Steuerpflichtigen, die negative Einkünfte erzielten, muss mittels des zu versteuernden Einkommens aller Steuerpflichtigen (zvE_j) der jeweiligen Klasse geschätzt werden. Die Anzahl der Steuerpflichtigen mit negativen Einkünften (g_j) und die Summe der negativen Einkünfte (G_j) je Klasse sind in den GdE-Tabellen enthalten. Daher ergeben sich bei der Gesamtsumme der negativen Einkünfte mit -0,17% beim Grundtarif und -0,35% beim Splittingtarif nur geringfügige Differenzen. Jedoch zeigt sich bei einem Vergleich der einzelnen Klassen, dass die Überleitung der nach den GdE-Klassen gegliederten negativen Einkünfte in eine nach zvE-Klassen gegliederte Tabelle zu hohen Abweichungen in den einzelnen Klassen führen kann. Dies resultiert daraus, dass bei der Gruppensimulation auf Grundlage von GdE-Tabellen die Verteilung des durchschnittlichen Betrags der negativen Einkünfte einer GdE-Klasse (G_j) entsprechend der Verteilung des zu versteuernden Einkommens aller Steuerpflichtigen dieser Klasse vorgenommen wird. Aus diesem Grund können auch das zu versteuernde Einkommen und das modifizierte zu versteuernde Einkommen in den einzelnen Klassen hohe Abweichungen aufweisen, wobei jedoch das nicht modifizierte zu versteuernde Einkommen ohne Berücksichtigung der Klasse der Steuerpflichtigen mit einem zu versteuernden Einkommen kleiner als eine DM relativ genau bestimmt werden kann (Abweichung 0,25% beim Grundtarif und -1,2% beim Splittingtarif).

Tabelle 4: Vergleich der Ausgangsdaten und des berechneten Steueraufkommens im Fall der Versagung des vertikalen Verlustausgleichs bei Mikrosimulation auf Basis einer Stichprobe und Gruppensimulation auf Basis von GdE-Tabellen der Einkommensteuerstatistik – Einkommensverteilung 1995

ZvE-Klasse Nr.	relative Abweichung zwischen den Simulationsergebnissen des StBA und denen des diskreten Gruppenmodells							
	negative Einkünfte		zu versteuerndes Einkommen		modifiziertes zu versteuerndes Einkommen		Berechnete ESt auf mod. zvE	
	Grd.-Tarif	**Sp.- Tarif**	**Grd.- Tarif**	**Sp.- Tarif**	**Grd.- Tarif**	**Sp.- Tarif**	**Grd.- Tarif**	**Sp.- Tarif**
1	-0,95%	-1,34%	-75,05%	-63,24%	73,73%	147,02%	189,31%	136,41%
2	-50,78%	-37,57%	-31,18%	-18,79%	-46,11%	-31,99%	-68,99%	-59,58%
3	-13,07%	-11,43%	0,06%	-1,08%	-7,96%	-6,68%	-29,48%	-26,05%
4	9,25%	-1,37%	13,27%	1,43%	11,13%	0,25%	-5,83%	-12,02%
5	17,25%	14,01%	19,80%	8,50%	18,53%	10,45%	6,81%	2,77%
6	29,06%	13,51%	18,99%	8,29%	23,66%	10,09%	11,29%	3,09%
7	8,07%	9,75%	6,03%	8,85%	6,91%	9,20%	-3,36%	2,21%
8	-7,58%	0,06%	-4,93%	2,19%	-5,88%	1,68%	-12,15%	-3,06%
9	-11,46%	5,50%	-3,57%	0,41%	-5,95%	1,39%	-11,73%	-1,15%
10	9,25%	16,26%	8,10%	6,99%	8,37%	8,56%	4,61%	6,37%
11	1,33%	58,41%	10,69%	23,48%	8,74%	29,04%	4,37%	28,36%
12	1,42%	40,82%	0,07%	14,03%	0,31%	18,20%	-1,98%	17,20%
13	22,78%	20,50%	16,16%	4,24%	17,33%	6,74%	14,63%	5,40%
14	5,67%	12,35%	11,61%	0,79%	10,48%	2,56%	7,42%	0,83%
15	8,27%	-5,75%	4,16%	-6,05%	4,87%	-6,00%	2,51%	-8,32%
16	-0,56%	-23,20%	1,82%	-16,78%	1,38%	-17,81%	-1,34%	-20,12%
17	28,67%	-30,56%	13,76%	-22,04%	16,32%	-23,39%	15,32%	-25,64%
18	21,9%	-55,51%	5,5%	-33,95%	8,49%	-37,79%	9,1%	-40,88%
19	36,01%	-32,40%	16,81%	-26,38%	20,69%	-27,50%	22,11%	-27,81%
20	-24,27%	1,15%	-17,82%	-0,42%	-19,11%	-0,15%	-19,82%	-0,25%
21	75,79%	9,33%	3,77%	0,89%	16,45%	2,18%	17,94%	2,33%
22	-2,36%	47,36%	-1,76%	0,51%	-1,85%	7,03%	-1,85%	7,42%
23	-68,43%	-40,02%	-15,47%	1,35%	-20,20%	-2,53%	-20,31%	-2,59%
gesamt	**-0,17%**	**-0,35%**	**19,30%**	**5,31%**	**26,87%**	**9,74%**	**2,80%**	**-1,21%**
gesamt ohne Klasse 1	**0,55%**	**0,18%**	**0,25%**	**-1,20%**	**21,85%**	**5,37%**	**-3,15%**	**-3,95%**
finanzielle Auswirkung							**16,47%**	**2,31%**
finanzielle Auswirkung ohne Klasse 1							**-5,09%**	**-8,79%**

Quelle: Statistisches Bundesamt, Wiesbaden; eigene Berechnungen.

Wie bereits bei der auf Grundlage von zvE-Tabellen durchgeführten Gruppensimulation zur Berechnung der Aufkommenswirkung des eingeschränkten Verlustausgleichs (Tabelle 2) zeigt sich auch bei den Gruppensimulationsrechnungen auf Grundlage von GdE-Tabellen, dass für die Klasse der Steuerpflichtigen mit einem zu versteuernden Einkommen kleiner als eine DM nur sehr ungenaue Simulationsergebnisse erzielt werden. Dies führt wiederum dazu, dass es bei der Bestimmung der gesamten finanziellen Auswirkungen des eingeschränkten Verlustausgleichs mittels des Gruppenmodells zu einer Überschätzung kommt. Bleibt die Klasse der Steuerpflichtigen, die kein positives zu versteuerndes Einkommen aufweisen, unberücksichtigt, ergeben sich für die finanziellen Auswirkungen mit einer Unterschätzung von 5,1% beim

Grundtarif und 8,8% beim Splittingtarif Abweichungen, die etwa in der Größenordnung der Differenzen bei der Verwendung von zvE-Tabellen liegen (Tabelle 2).

V. Zusammenfassung

Mit den oben dargestellten Vergleichen der Ergebnisse der Mikrosimulation mit denen der Simulation mittels des diskreten Gruppenmodells auf Grundlage unterschiedlich klassierter Daten sollten die möglichen Auswirkungen der vereinfachenden Vorgehensweise beim Gruppenmodell sowie des Informationsverlusts durch die aggregierten und unvollständigen Daten anhand von Beispielen aufgezeigt und quantifiziert werden. Eine exakte formale Bestimmung der möglichen Abweichungen der Simulationsergebnisse des Gruppenmodells von denen einer Mikrosimulation war nicht Gegenstand dieser Untersuchung und würde den Rahmen der Arbeit überschreiten. Die in den Beispielen ermittelten Differenzen stellen keine allgemein gültigen Werte dar. Sie sollen lediglich einen Eindruck von den Größenordnungen der möglichen Ungenauigkeiten einer Gruppensimulation vermitteln.

Zusammenfassend lässt sich feststellen, dass mit dem hier vorgestellten Gruppensimulationsmodell bei der einfachen Steuertarifsimulation sehr genaue Ergebnisse erzielt werden können. Bei der Bestimmung der finanziellen Auswirkungen von modifizierten steuerlichen Bemessungsgrundlagen, insbesondere bei Veränderungen, die bei jedem Steuerpflichtigen individuell verschieden hoch sein können, nehmen – vor allem wegen der Klasse der Steuerpflichtigen, die ein zu versteuerndes Einkommen kleiner als eine DM aufweisen – die Differenzen zu den Ergebnissen einer Mikrosimulation zu. Bleibt die Klasse der Steuerpflichtigen mit einem zu versteuernden Einkommen kleiner als eine DM unberücksichtigt, kommt es bei der Simulation der finanziellen Auswirkungen einer Bemessungsgrundlagenerhöhung unter Anwendung des Gruppenmodells bei einem progressiven Steuertarif zu einer systematischen Unterschätzung, da in Ermangelung disaggregierter Daten für die Modifikation der Bemessungsgrundlage innerhalb einer Klasse für jeden Steuerpflichtigen der empirische Klassendurchschnitt des Abzugsbetrags hinzugerechnet wird. Liegen der Gruppensimulation Daten zu Grunde, die nicht nach dem zu versteuernden Einkommen, sondern nach dem Gesamtbetrag der Einkünfte gegliedert sind, so führt dies tendenziell sowohl beim Gesamtergebnis als auch bei den einzelnen Klassen zu höheren Abweichungen von den Mikrosimulationsergebnissen.

Daraus könnte nun geschlussfolgert werden, dass – soweit dies möglich ist – stets eine Mikrosimulation durchgeführt werden sollte. Dabei muss jedoch berücksichtigt werden, dass auch die Berechnungen des Statistischen Bundesamtes zu sehr großen Abweichungen führen können, soweit sie auf Grundlage einer Stichprobe durchgeführt werden. Dies ist darauf zurückzuführen, dass die geschichtete Stichprobe[90] für Sachverhalte mit einer geringen Häufigkeit nicht immer repräsentativ ist.[91] in diesem Fall sind die durch das Gruppenmodell und die aggregierte Datenbasis bedingten möglichen Fehler wesentlich geringer als

[90] Zur Schichtung der Stichprobe der Einkommensteuerstatistik vgl. Zwick, WiSta 1998, S. 566 (570–571).

[91] Dies dürfte auch für die Datenbasis des Potsdamer Modells gelten. *Bork* (Fn. 4), S. 95 – 96, schlussfolgert aus dem Vergleich der Daten der Einkommensteuerstatistik mit denen der Volkswirtschaftlichen Gesamtrechnung, dass wegen der Abweichungen – so erreichen z.B. die in der Steuerstatistik 1992 ausgewiesenen Einkünfte aus nichtselbständiger Arbeit nur ca. 90% der Bruttolohn- und Gehaltssumme (vgl. *Bork* (Fn. 4), S. 110) – die Qualität der Daten der Lohn- und Einkommensteuerstatistik erheblich begrenzt wäre. Aber hinsichtlich der Repräsentativität der steuerrelevanten Daten dürfte die geschichtete Stichprobe des Statistischen Bundesamtes mit ca. drei Mio. Datensätzen (vgl. *Zwick*, WiSta 1998, S. 566 (570 – 573), und *Zwick* (Fn. 24) , S. 261 – 264) der Datenbasis des Potsdamer Modells mit 51.536 Steuerpflichtigen (vgl. *Bork* (Fn. 4), S. 89), die entsprechend hochgerechnet werden, überlegen sein.

die bei einer Mikrosimulation, da die Gruppensimulation auf Grundlage der Werte der Grundgesamtheit der Einkommensteuerstatistik erfolgt.

Hinsichtlich des Ziels die Aufkommenswirkungen alternativer Zustände, insbesondere der finanziellen Auswirkungen bestimmter steuerlicher Regelungen, zu ermitteln und zu analysieren, stellt das hier vorgestellte Gruppensimulationsmodell einen guten Kompromiss zwischen dem Anliegen einer möglichst vollständigen modellhaften Erfassung der komplexen Realität und der erzielbaren Genauigkeit bei eingeschränkt verfügbaren Ressourcen und Daten dar. Die Aufkommenswirkungen relativ einfach zu modellierender Sachverhalte können mit dem diskreten Gruppensimulationsmodell bestimmt werden. Zudem ist bei bestimmten Fragestellungen, bei denen die Stichprobe nicht repräsentativ ist, nicht nur aus Kostengründen, sondern auch hinsichtlich der Verlässlichkeit der Ergebnisse, die Gruppensimulation einer Mikrosimulation auf Grundlage einer Stichprobe vorzuziehen. Für die Untersuchung von alten Datenbeständen wird die Gruppensimulation wegen der nicht verfügbaren Einzeldaten weiterhin von Bedeutung sein.