
WISTA

Wirtschaft und Statistik

Kathrin Gebers Philip Graze	Statistische Datengewinnung durch die Nutzung geografischer Informationen
Christian Blaudow Daniel Seeger	Fortschritte beim Einsatz von Web Scraping in der amtlichen Verbraucherpreisstatistik – ein Werkstattbericht
Maren Köhlmann	Datenergänzung der Strukturerhebung Energie mithilfe eines robusten Regressionsmodells
Jana Emmenegger Tobias Godec Prof. Dr. Markus Zwick	Simulationsprojekt zu Dienstunfällen von Beamtinnen und Beamten
Nelli Krüger	Maßnahmen zur Steigerung der Teilnahmebereitschaft an Haushaltserhebungen der amtlichen Statistik
Henrik Neumann Kristina Trusheim	Harmonisierung der Umsatzsteuer-Meldefristen

4 | 2019

ABKÜRZUNGEN

D	Durchschnitt (bei nicht addierfähigen Größen)
Vj	Vierteljahr
Hj	Halbjahr
a. n. g.	anderweitig nicht genannt
o. a. S.	ohne ausgeprägten Schwerpunkt
Mill.	Million
Mrd.	Milliarde

ZEICHENERKLÄRUNG

–	nichts vorhanden
0	weniger als die Hälfte von 1 in der letzten besetzten Stelle, jedoch mehr als nichts
.	Zahlenwert unbekannt oder geheim zu halten
...	Angabe fällt später an
X	Tabellenfach gesperrt, weil Aussage nicht sinnvoll
I oder –	grundsätzliche Änderung innerhalb einer Reihe, die den zeitlichen Vergleich beeinträchtigt
/	keine Angaben, da Zahlenwert nicht sicher genug
()	Aussagewert eingeschränkt, da der Zahlenwert statistisch relativ unsicher ist

Abweichungen in den Summen ergeben sich durch Runden der Zahlen.

Tiefer gehende Internet-Verlinkungen sind in der Online-Ausgabe hinterlegt.

INHALT

3	Editorial
4	Kennzahlen
6	Kurznachrichten
11	Kathrin Gebers, Philip Graze Statistische Datengewinnung durch die Nutzung geografischer Informationen <i>Collecting statistical data by using geographical information</i>
19	Christian Blaudow, Daniel Seeger Fortschritte beim Einsatz von Web Scraping in der amtlichen Verbraucherpreisstatistik – ein Werkstattbericht <i>Progress made in the use of web scraping for official consumer price statistics – A workshop report</i>
31	Maren Köhlmann Datenergänzung der Strukturhebung Energie mithilfe eines robusten Regressionsmodells <i>Supplementing data in the structural survey of energy supply, using a robust regression model</i>
44	Jana Emmenegger, Tobias Godec, Prof. Dr. Markus Zwick Simulationsprojekt zu Dienstunfällen von Beamtinnen und Beamten <i>Simulation project on work accidents of public officials</i>

INHALT

- | | |
|----|---|
| 56 | Nelli Krüger
Maßnahmen zur Steigerung der Teilnahmebereitschaft an Haushalts-
erhebungen der amtlichen Statistik
<i>Measures for increasing the willingness to participate in household
surveys of official statistics</i> |
| 67 | Henrik Neumann, Kristina Trusheim
Harmonisierung der Umsatzsteuer-Meldefristen
<i>Harmonisation of turnover tax reporting deadlines</i> |

EDITORIAL

Dr. Georg Thiel



LIEBE LESERIN, LIEBER LESER,

mit der vorliegenden Ausgabe wechselt die Schriftleitung von WISTA: Die langjährige Leiterin der Abteilung „Bevölkerung, Finanzen und Steuern“, Dr. Sabine Bechtold, hatte diese Aufgabe seit Herbst 2017 mit großem Engagement und fachlicher Expertise wahrgenommen. Ich danke ihr für ihren Einsatz für WISTA und hoffe, sie auch im Ruhestand weiterhin zu unserer Leserschaft zählen zu können.

Ihre Nachfolge als Schriftleiter tritt Dr. Daniel Vorgrimler an, der neue Leiter der Abteilung „Strategie und Planung, Internationale Beziehungen, Forschung und Kommunikation“. Seine bisherigen Stationen im Statistischen Bundesamt führten den Volkswirt von der faktischen Anonymisierung wirtschaftsstatistischer Einzeldaten über die Einkommensteuerstatistik und die Bürokratiekostenmessung zuletzt in die Gruppe „Industrie, Bau, Energie“. Als Abteilungsleiter ist es nun sein Ziel, die Relevanz amtlicher Daten für Politik und Gesellschaft weiter zu steigern. Seine Funktion als Schriftleiter von WISTA wird ihn dabei unterstützen. Dafür wünsche ich ihm gutes Gelingen!

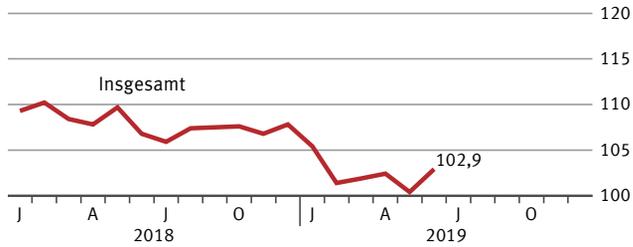
In der vorliegenden Ausgabe erwarten Sie natürlich wieder interessante Artikel rund um die methodischen Weiterentwicklungen in der amtlichen Statistik. Einen Beitrag möchte ich hervorheben: In einem Projekt haben das Hessische Statistische Landesamt und das Statistische Bundesamt durch die Integration statistischer und geografischer Daten neue Informationen über hessische Grundschulstandorte erschlossen. Nach diesen experimentellen Daten wäre für rund 90% der hessischen Familien mit mindestens einem Kind im grundschulpflichtigen Alter die nächstgelegene Grundschule höchstens zwei Kilometer entfernt.

Ich wünsche Ihnen eine spannende Lektüre.

Präsident des Statistischen Bundesamtes

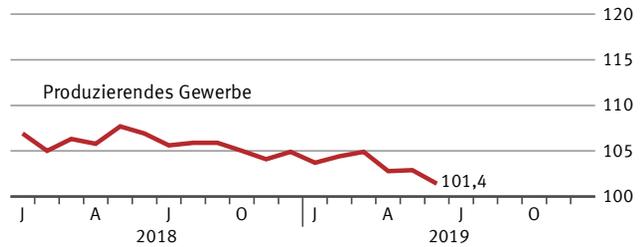
Kennzahlen

Auftragseingang im Verarbeitenden Gewerbe Volumenindex 2015 = 100



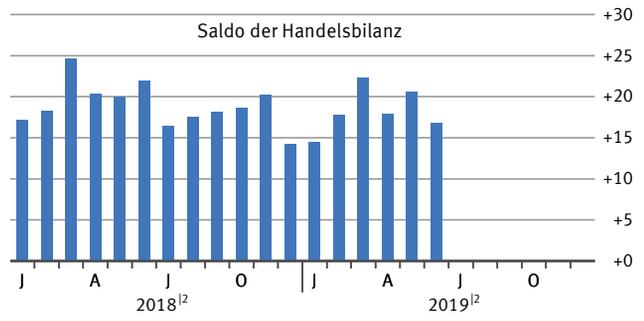
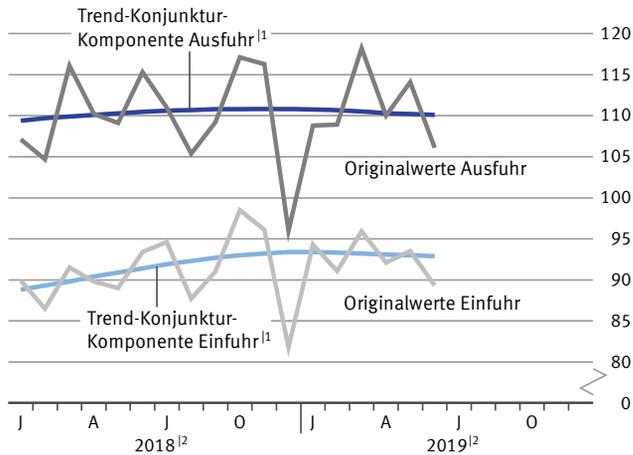
Kalender- und saisonbereinigter Wert nach dem Verfahren X13 JDemetra+. – Vorläufiges Ergebnis.

Produktion im Produzierenden Gewerbe Index 2015 = 100



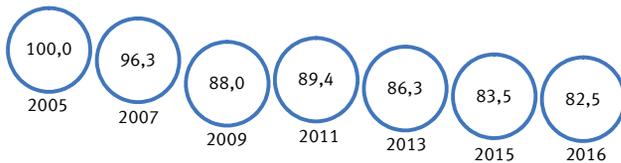
Kalender- und saisonbereinigter Wert nach dem Verfahren X13 JDemetra+. – Vorläufiges Ergebnis.

Außenhandel in Mrd. EUR

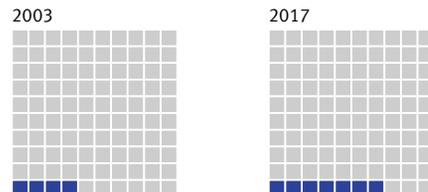


1 Berechnung nach dem Berliner Verfahren, Version 4.1 (BV 4.1).
2 Vorläufige Ergebnisse.

Luftbelastung Emissionen von Luftschadstoffen, 2005 = 100



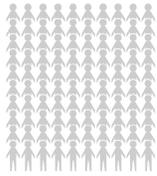
Ökologische Landwirtschaft Anteil an der landwirtschaftlich genutzten Fläche in %



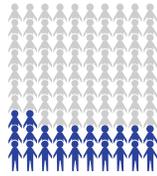
Kennzahlen

Bevölkerung Stand: 31.12.2018

Insgesamt
83 Millionen

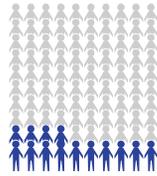


über 65 Jahre



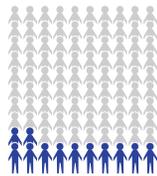
22%

unter 15 Jahre



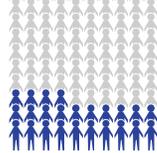
14%

Ausländeranteil



12%

Migrationshintergrund
(31.12.2017)



24%

Verbraucherpreisindex 2015 = 100

2018

Januar	102,0
Februar	102,3
März	102,9
April	103,1
Mai	103,9
Juni	104,0
Juli	104,4
August	104,5
September	104,7
Oktober	104,9
November	104,2
Dezember	104,2

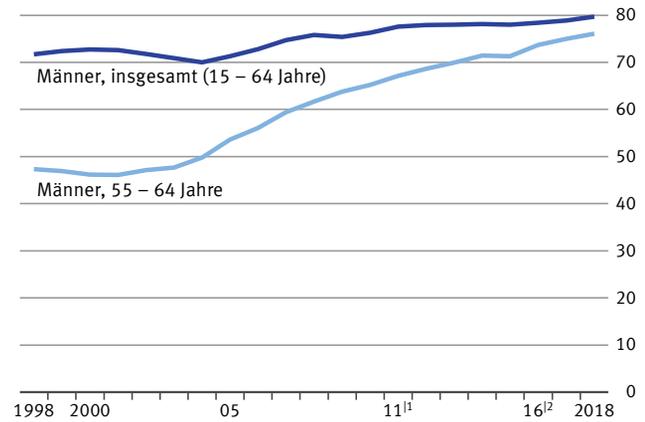
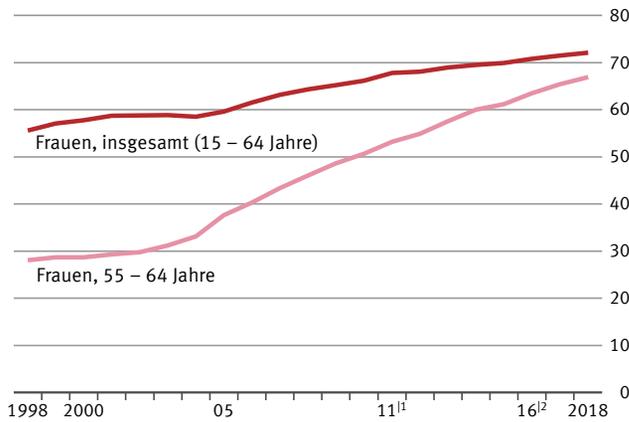
2019

Januar	103,4
Februar	103,8
März	104,2
April	105,2
Mai	105,4
Juni	105,7

Veränderung
zum Vorjahresmonat

105,7 ↑ **1,6 %**

Erwerbstätigenquote Anteile in %

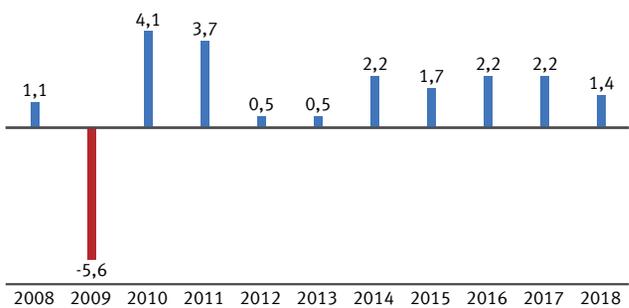


Bis 2004: Ergebnisse für eine feste Berichtswoche im Frühjahr; ab 2005: Jahresdurchschnittsergebnisse sowie geänderte Erhebungs- und Hochrechnungsverfahren.

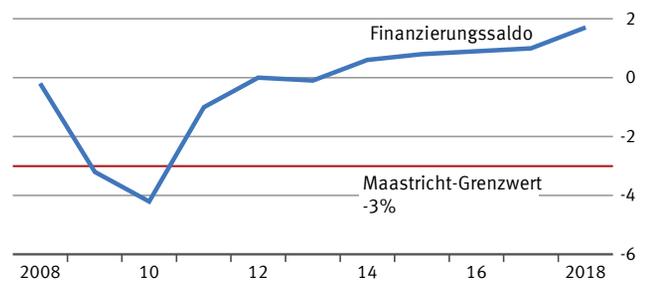
1 Ab 2011: Hochrechnung anhand von Bevölkerungseckwerten auf Basis des Zensus 2011.

2 Ab 2016: aktualisierte Stichprobe auf Grundlage des Zensus 2011. Zeitreihenvergleiche nur eingeschränkt möglich.

Bruttoinlandsprodukt preisbereinigt, Veränderung gegenüber dem Vorjahr in %



Finanzierungssaldo des Staates in % des Bruttoinlandsprodukts



KURZNACHRICHTEN

IN EIGENER SACHE

Neue Website zum Zensus 2021

Im Jahr 2021 findet in Deutschland wieder ein Zensus statt – wie bereits 2011 auf Daten aus Verwaltungsregistern gestützt, mit einer Stichprobe ergänzt und mit einer Gebäude- und Wohnungszählung kombiniert. Mit dieser statistischen Erhebung wird ermittelt, wie viele Menschen in Deutschland leben, wie sie wohnen und arbeiten. Viele Entscheidungen in Bund, Ländern und Gemeinden beruhen auf Bevölkerungs- und Wohnungszahlen.

Die neue Website www.zensus2021.de der Statistischen Ämter des Bundes und der Länder informiert allgemein und verständlich über den Zensus. Das Design der Website ist für mobile Endgeräte optimiert und wird die Befragten beim Zensus 2021 zum Fragebogen führen, den sie bequem am Smartphone ausfüllen können. Im September 2019 wird mithilfe der neuen Website ein quantitativer Pretest der Gebäude- und Wohnungszählung durchgeführt. Die Teilnehmerinnen und Teilnehmer können sich dann über einen direkten Zugang über die Startseite zum Onlinefragebogen anmelden.

↳ www.zensus2021.de

Neues auf www.destatis.de

Das Onlineangebot des Statistischen Bundesamtes wird ständig optimiert und weiterentwickelt. Nach dem umfassenden Relaunch der Destatis-Website im März 2019 stand allen fremdsprachigen Nutzerinnen und Nutzern eine provisorische englische Startseite zur Verfügung, nun gibt es wieder umfassende Informationen in englischer Sprache.

Eine weitere Neuerung sind interaktive Diagramme, die die Ergebnisse der Kalender- und Saisonbereinigung anschaulich darstellen. Die Werte für diese Diagramme werden monatlich oder quartalsweise mithilfe automatisierter Prozesse aktualisiert, um die laufende Pflege so effizient wie möglich zu gestalten. Zur Visualisierung wird Highcharts verwendet, eine Software für interaktive Grafiken.

Mithilfe von Highcharts werden verschiedene Typen von Diagrammen erstellt, die sich je nach Endgerät responsiv verhalten und Nutzerinnen und Nutzern Funktionen wie das Herunterladen statischer Grafiken oder der zugrunde liegenden Daten ermöglichen.

↳ www.destatis.de

Deutsche Online-Plattform zur Entwicklung globaler Nachhaltigkeitsziele

Die globalen Nachhaltigkeitsziele wurden im Jahr 2015 durch die Generalversammlung der Vereinten Nationen verabschiedet. Sie streben die Verbesserung der Lebensverhältnisse aller Menschen sowie den Schutz des Planeten Erde an. Um Fortschritte in diesem Bestreben zu messen, hat eine internationale Arbeitsgruppe unter Beteiligung des Statistischen Bundesamtes ein Set von 244 Indikatoren entwickelt. In Deutschland ist das Statistische Bundesamt dafür verantwortlich, die Daten zu diesen Indikatoren zu veröffentlichen sowie die Datenlieferung an internationale Organisationen zu koordinieren.

Das Statistische Bundesamt stellt Daten zu den 17 globalen Nachhaltigkeitszielen der Vereinten Nationen

(Sustainable Development Goals, SDGs) seit Kurzem auf einer neuen Online-Plattform bereit. Die nationale Berichtsplattform stellt deutsche Daten und Metadaten zur globalen nachhaltigen Entwicklung in interaktiver Form zur Verfügung. Sie zeigen den Stand der Entwicklung in Deutschland bezogen auf die Nachhaltigkeitsziele der Vereinten Nationen.

↳ <https://sustainabledevelopment-germany.github.io/>

AUS ALLER WELT

Committee on Statistics and Statistical Policy

Der Ausschuss für Statistik und Statistikpolitik (Committee on Statistics and Statistical Policy – CSSP) der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) behandelte bei seiner Sitzung im Juni 2019 unter anderem folgende Punkte:

- › Die Arbeiten im Kontext der Messung der digitalen Wirtschaft sowie die vorgesehene Roadmap zeigen gute Fortschritte, so werden zum Beispiel im Rahmen des Projekts „Measuring the Digital Transformation“ bereits 200 relevante Indikatoren bereitgestellt. Insbesondere der Messung von neuartigen Arbeitsverhältnissen (zum Beispiel „platform workers“) wurde eine besondere Priorität eingeräumt. Um weitere Informationen zur Messung der digitalen Wirtschaft zu gewinnen, verständigte sich der CSSP auf die Maßnahmen „Digitale Wirtschaft in den Wirtschaftsstatistiken sichtbar machen“, „Wirtschaftliche Auswirkungen der digitalen Transformation verstehen“, „Messung des Well-being im digitalen Zeitalter“ und „Neue Ansätze der Datensammlung“.
- › Die OECD stellte ihre Arbeiten im Rahmen der Umweltökonomischen Gesamtrechnungen und der Umsetzung des System of Environmental-Economic Accounting (SEEA) vor. Unter anderem hat die OECD Indikatoren entwickelt, die Erdbeobachtungsquellen in Bereichen nutzen, in denen bislang keine harmonisierten Informationen zur Verfügung standen (zum Beispiel Umwandlungen in Ackerland und von Ackerland zu versiegelten Flächen sowie städtebauliche

Erweiterungen). Hierbei wird das Thema „Geography of Well-being“ als neuer Trend gesehen, indem Umweltstatistiken und ökonomische Statistiken miteinander verknüpft werden.

- › Zum Thema „Wohlbefinden und Nachhaltigkeit“ wurden aktuelle Erfahrungen ausgetauscht, unter anderem berichtete die OECD zum Indikatorenset „How’s Life“ sowie zur Entwicklung von Dashboards. Generell soll angestrebt werden, die Diskussion zu Nachhaltigkeit und „well-being“ in einen Zusammenhang zu bringen, da die Überschneidungsbereiche sehr groß sind. Neben zusammengeführten Indikatoren sollen auch leicht verständliche Daten angeboten werden.

AUS EUROPA

66. CES-Plenarsitzung

Im Mittelpunkt der diesjährigen Tagung der Konferenz Europäischer Statistiker (CES) im Juni 2019 standen folgende Seminarthemen:

- › New data sources – accessibility and use
Dieses Seminar befasste sich mit den Möglichkeiten und Herausforderungen, denen Statistikämter bei der Nutzung neuer Datenquellen begegnen. Auch die Fähigkeiten von Beschäftigten, die notwendig sind, um neue Datenquellen erfolgreich zu nutzen, wurden thematisiert. Fazit war, dass noch weit mehr Koordinierung und ein stärker abgestimmtes Vorgehen erforderlich sind als bisher. In allen Ländern bestehen ähnliche Probleme, was die Abdeckung der Daten, die entstehenden Kosten und die rechtlichen Rahmenbedingungen anbelangt.
- › Emerging role of national offices as offices for statistics and data
Das zweite Seminar thematisierte die zunehmende Bedeutung der Statistikämter als mögliche „data stewards“ (Datenmanager) für die nationalen Regierungen. Die Statistik muss hierbei eine führende Position übernehmen, insbesondere hinsichtlich einer Vorgabe von Standards im Spannungsfeld zwischen experimentellen und amtlichen Statistiken. Als besonders wichtig wurde hier der Aufbau von Partnerschaften angesehen.

AUS DEM INLAND

Ergebnisse der Kommission „Gleichwertige Lebensverhältnisse“ und Deutschlandatlas

Das Bundeskabinett hat im Juli 2019 die Schlussfolgerungen aus der Arbeit der Kommission „Gleichwertige Lebensverhältnisse“ zur Kenntnis genommen und zwölf konkrete Maßnahmen des Bundes zur Umsetzung beschlossen. Die Schlussfolgerungen „Unser Plan für Deutschland – Gleichwertige Lebensverhältnisse überall“ fassen die Ergebnisse der Arbeit der Kommission „Gleichwertige Lebensverhältnisse“ aus Sicht des Bundes zusammen.

Zur Unterstützung der Kommission „Gleichwertige Lebensverhältnisse“ wurde im Statistischen Bundesamt eine Servicestelle eingerichtet. Diese hilft den Arbeitsgruppen der Kommission bei datenbezogenen Fragestellungen, recherchiert verfügbare Datenquellen und bereitet Daten nutzerorientiert auf.

↳ www.bmi.bund.de

Neben den Schlussfolgerungen wurde auch der „Deutschlandatlas“ vorgestellt. Dieser stellt eine Bestandsaufnahme quer durch alle Lebensbereiche dar. Der Deutschlandatlas visualisiert in 56 Karten ein Themenspektrum von der Flächennutzung und Bevölkerungsstruktur über die Kommunalfinanzen bis hin zu Arbeit, Bildung und Einkommensverteilung sowie sozialen und kulturellen Aspekten.

↳ <https://heimat.bund.de/atlas/>

Neuer nationaler Ausschuss für Produktivität

Die Bundesregierung hat dem Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung mit Wirkung zum 1. August 2019 die Aufgabe des nationalen Ausschusses für Produktivität übertragen. Damit setzt sie eine Empfehlung des Rates der Europäischen Union vom 20. September 2016 um.

Die nationalen Ausschüsse in den jeweiligen Mitgliedstaaten sollen unabhängige Analysen zu den Entwicklungen im Bereich der Produktivität und Wettbewerbsfähigkeit veröffentlichen und dabei nationale Besonderheiten berücksichtigen. Dazu zählen etwa Diagnosen zu den langfristigen Antriebsfaktoren der wirtschaftlichen Entwicklung und den Voraussetzungen, wie Produktivität und Wettbewerbsfähigkeit gesteigert werden können. Ein breiter Konsultationsprozess auf nationaler Ebene sowie der regelmäßige Austausch und eine intensive Zusammenarbeit auf europäischer Ebene sollen in die Analysen einfließen.

Die neue Aufgabe wird eine weitere Intensivierung der Zusammenarbeit mit den verschiedenen Bereichen des Statistischen Bundesamtes mit sich bringen: Ein Großteil der für die Analysen notwendigen statistischen Datengrundlagen stammt beispielsweise aus den Volkswirtschaftlichen Gesamtrechnungen und den Unternehmensstatistiken.

VERANSTALTUNGEN

Konferenz für Sozial- und Wirtschaftsdaten

Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) beruft alle drei Jahre eine Konferenz für Sozial- und Wirtschaftsdaten ein. Die 8. Konferenz (8|KSWD) findet am 2. und 3. März 2020 im Konferenzzentrum der Heinrich-Böll-Stiftung in Berlin statt mit dem Thema

Gesellschaft braucht Wissenschaft.
Wissenschaft braucht Daten.

Weitere Informationen sowie ein Call for Posters sind der Website des RatSWD zu entnehmen:

↳ www.ratswd.de/kswd

Konferenz „Messung der Preise“

Anfang Juni 2019 fand in Jena die 23. Konferenz „Messung der Preise“ statt. Ausgerichtet wurde die Veranstaltung vom Amt für Statistik Berlin-Brandenburg und vom Thüringer Landesamt für Statistik als Gastgeber.

Die Konferenz bietet jährlich eine Plattform für den Austausch zwischen Forschung, Statistikproduzenten und wichtigen Nutzergruppen zu Fragen der Preisstatistik.

Die Vorträge behandelten unter anderem Analysen der Revisionsergebnisse des Verbraucherpreisindex und des harmonisierten Verbraucherpreisindex, indextheoretische Fragestellungen, Nutzung und Ausblicke zum Umgang mit Scannerdaten sowie Informationen zu Forschungsvorhaben. Das Statistische Bundesamt beteiligte sich mit Vorträgen zu den Themen Revision der Verbraucherpreise 2015 = 100, Experimenteller Preisindex für Pauschalreisen auf Basis von Transaktionsdaten, Mieten – neues Wägungsschema, tiefer gegliederte Ergebnisse sowie Scannerdaten – Sachstand und Ausblick.

↳ www.statistik-berlin-brandenburg.de

Arbeitsmarkt zwischen Digitalisierung und Wertewandel

Anfang Juni 2019 trafen sich Expertinnen und Experten aus der Arbeitsmarktforschung, der amtlichen Statistik und der Arbeitsmarktpolitik zur wissenschaftlichen Tagung „Arbeitsmarkt zwischen Digitalisierung und Wertewandel – neue Indikatoren zur Qualität der Arbeit“ im Statistischen Bundesamt.

In welchem Spannungsfeld stehen die technisch bedingte Veränderung der Arbeitswelt und die sich wandelnden Ansprüche der Beschäftigten? Das war die Frage, die knapp 70 Teilnehmerinnen und Teilnehmer aus Wissenschaft und Politik, Verbänden, Bund, Ländern und von internationalen Institutionen diskutierten. Unter anderem ging es darum, welche Auswirkungen die Digitalisierung auf den Arbeitsmarkt und vor allem auf die Anforderungen an die Beschäftigten hat, wie sich dies auf die Qualität der Arbeit auswirkt und mit welchen Indikatoren diese Entwicklung national wie international abgebildet werden kann. Die vorgestellten wissenschaftlichen Studien beleuchteten diese Zusammenhänge anhand etablierter, aber auch neu erschlossener Datenquellen.

↳ www.destatis.de

Qualität bei zusammengeführten Daten – Befragungsdaten, Administrative Daten, Neue digitale Daten: Miteinander besser?

Zu diesem Thema veranstaltete das Statistische Bundesamt gemeinsam mit dem ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. und der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (ASI) am 27. und 28. Juni 2019 in Wiesbaden eine wissenschaftliche Fachtagung. Die Kurzfassungen und Folienpräsentationen stehen auf der Website des Statistischen Bundesamtes zur Verfügung; die Langfassungen der Beiträge werden später in einem Tagungsband erscheinen.

↳ www.destatis.de

NEUERSCHEINUNGEN

Sondergutachten des Sachverständigenrats

Der Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung wurde von der Bundesregierung beauftragt, ein Sondergutachten zur Bepreisung von Kohlenstoffdioxid (CO₂) zu erstellen. Im Sondergutachten „Aufbruch zu einer neuen Klimapolitik“ diskutiert er Reformoptionen für die Klimapolitik.

↳ www.sachverstaendigenrat-wirtschaft.de

STATISTISCHE DATENGEWINNUNG DURCH DIE NUTZUNG GEOGRAFISCHER INFORMATIONEN

Kathrin Gebers, Philip Graze

↳ **Schlüsselwörter:** Digitalisierung – Datenintegration – Geodaten – Georeferenzierung – Erreichbarkeitsanalyse

ZUSAMMENFASSUNG

In der amtlichen Statistik werden durch die Integration statistischer und geografischer Daten Informationen nutzbar gemacht, die bisher nicht erschlossen werden konnten. Das daraus entstehende Wissen kommt in den unterschiedlichsten Lebensbereichen (beispielsweise Wirtschaft, Gesellschaft, Umwelt) zur Anwendung. Die Statistischen Ämter des Bundes und der Länder stellen statistische Informationen bereit, die zur Willensbildung und in Entscheidungsprozessen in einer demokratischen Gesellschaft notwendig sind. Diese Aufgabe wird mit den neu erschlossenen integrierten Daten auf eine neue Ebene gebracht.

Modellhaft veranschaulicht wird dies am konkreten Beispiel der Erweiterung des hessischen Mikrozensus-Datensatzes um die berechnete Erreichbarkeit von Grundschulen.

↳ **Keywords:** digitalisation – data integration – geodata – geo-referencing – accessibility analysis

ABSTRACT

In official statistics, the integration of statistical and geographical information delivers insights that previously could not be accessed. The resulting knowledge is used in the most diverse areas of life (e. g. economy, society, and environment). The statistical offices of the Federation and the Länder provide statistical information necessary for developing informed opinions and decision-making in a democratic society. This task is taken to a new level by such integrated data that have become accessible.

This is illustrated by the concrete example of extending the microcensus data set for Hessen to include the calculated accessibility of primary schools.



Kathrin Gebers

ist M.Sc. in Economics und als Referentin im Referat „Methoden der Datenanalyse“ des Statistischen Bundesamtes tätig. Schwerpunktmäßig beschäftigt sie sich mit der Integration statistischer und geografischer Daten sowie deren Analyse.



Philip Graze

studierte Wirtschafts- und Sozialgeografie. Er war Datenanalyst für das Bundesinstitut für Bevölkerungsforschung und arbeitet seit 2009 für das Hessische Statistische Landesamt. Dort verantwortet er die Themen Geoinformation und Veröffentlichungen.

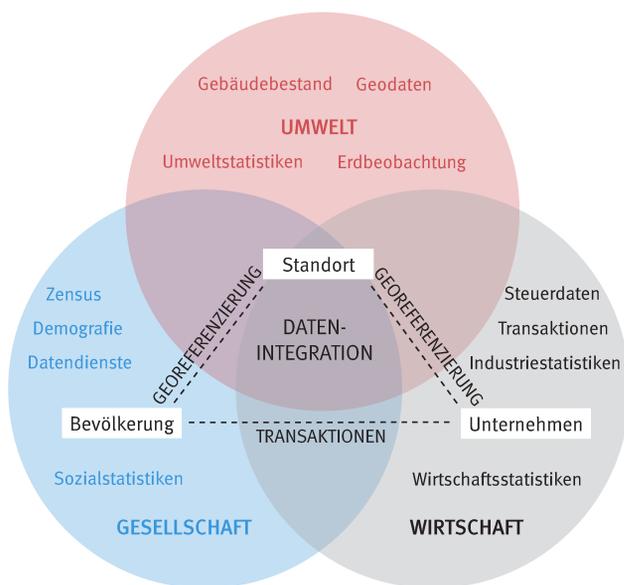
1

Einleitung

Mit ihrem Bezug zu einem bestimmten Standort oder Gebiet sind Geodaten ein wichtiger strategischer Baustein der Digitalisierung. Über den Raumbezug lassen sich Informationen verschiedenster Quellen verknüpfen. Das daraus entstehende Wissen kann in unterschiedlichen Lebensbereichen, wie Wirtschaft, Gesellschaft oder Umwelt, genutzt werden. [↪ Grafik 1](#)

Grafik 1

Raumbezug als Verknüpfungspunkt von Wirtschaft, Gesellschaft und Umwelt



Quelle: Sitzungsunterlage des „Committee of Experts on Global Geospatial Information Management“ der Vereinten Nationen 2019 - 01 - 0518

Somit kann der Auftrag der Statistischen Ämter des Bundes und der Länder, statistische Informationen bereitzustellen, die zur Willensbildung und in Entscheidungsprozessen in einer demokratischen Gesellschaft notwendig sind, deutlich umfassender erfüllt werden als bisher. Durch die Mehrfachnutzung und Integration von nicht zentral vorliegenden Datenbeständen wird eine neue, erweiterte Grundlage für evidenzbasierte Entscheidungen mit Regionalbezug geschaffen (van Halderen und andere, 2016, hier: Seite 463).

Im Folgenden wird gezeigt, wie in der amtlichen Statistik Informationen, die bisher nicht erschlossen werden konnten, durch die Integration statistischer und geografischer Daten nutzbar gemacht werden. Dies geschieht durch die Kombination und Auswertung vorhandener Datenbestände ohne zusätzliche Belastung der Auskunftspflichtigen.

Die Verwendung neuer Erhebungs- und Analysemethoden ist auch Bestandteil der Digitalen Agenda der Statistischen Ämter des Bundes und der Länder beziehungsweise Teil des digitalen Transformationsprozesses: *«In der amtlichen Statistik streben wir vom Grundsatz eine digitale Transformation an, also vorhandene Prozesse nicht einfach weiter zu automatisieren, sondern – soweit erforderlich und möglich – unsere Geschäftsprozesse um neue Wege und Möglichkeiten zu erweitern.»* (Statistische Ämter des Bundes und der Länder, 2018, hier: Seite 6).

2

Geografische Raster als Analysebasis

Um Datenbestände räumlich verknüpfen zu können, braucht es dieselbe räumliche Basiseinheit in verschiedenen Datenbeständen. Eine Möglichkeit bieten geografische Raster. Geografische Raster sind räumliche geometrische Bezugseinheiten, die in ihrer Größe und Form einheitlich sind. Diese haben gegenüber den administrativen Gebietseinheiten (Landkreise, Gemeinden und so weiter) den Vorteil, dass räumliche und zeitliche Vergleiche oder statistische Analysen im geografischen Kontext problemlos operationalisierbar sind.

Vor allem aber können geografische Raster genutzt werden, um bestehende Statistiken um zusätzliche Merkmale zu erweitern (Scholze, 2015, hier: Seite 211). Seit der Änderung des Bundesstatistikgesetzes vom 1. August 2013 wird durch §10 Absätze 2 und 3 das dauerhafte Führen einer quadratischen geografischen Rasterzelle von 1 Hektar und der Weite 100 m x 100 m an den Erhebungseinheiten ermöglicht.

Rasterzellen sind Raumeinheiten, denen Attribute zugeschrieben werden können. Dies bietet großes Potenzial, da kleinräumige Informationen hoher Auflösung aus unterschiedlichen Quellen, wie Neue Digitale Daten

↘ Auszug aus § 10 Bundesstatistikgesetz – Erhebungs- und Hilfsmerkmale

(2) Der Name der Gemeinde, die Blockseite und die geografische Gitterzelle dürfen für die regionale Zuordnung der Erhebungsmerkmale genutzt werden. Die übrigen Teile der Anschrift dürfen für die Zuordnung zu Blockseiten und geografischen Gitterzellen für einen Zeitraum von bis zu vier Jahren nach Abschluss der jeweiligen Erhebung genutzt werden. Besondere Regelungen in einer Bundesstatistik anordnenden Rechtsvorschrift bleiben unberührt.

(3) Blockseite ist innerhalb eines Gemeindegebiets die Seite mit gleicher Straßenbezeichnung von der durch Straßeneinmündungen oder vergleichbare Begrenzungen umschlossenen Fläche. Eine geografische Gitterzelle ist eine Gebietseinheit, die bezogen auf eine vorgegebene Kartenprojektion quadratisch ist und mindestens 1 Hektar groß ist.

oder administrative Daten, technisch einfach miteinander verknüpft werden können.

Der Fokus dieses Artikels und der Integration statistischer und geografischer Informationen im Allgemeinen liegt nicht auf der kartografischen Darstellung georeferenzierter Informationen, sondern vielmehr auf den Möglichkeiten und Ansätzen für erweiterte Analysen. Daher wird die thematische Verschneidung von Informationen über den Raumbezug vorgestellt und es werden beispielhaft Ergebnisse der dadurch entstandenen Analysemöglichkeiten gezeigt. Insbesondere die Sinnhaftigkeit und der Mehrwert, der durch die Georeferenzierung von Stichproben erzeugt werden kann, stehen somit im Vordergrund.

3

Integration statistischer und geografischer Daten

3.1 Fragestellung

Wie stellt sich die durchschnittliche Wegstrecke für hessische Familien mit mindestens einem grundschulpflichtigen Kind im Alter von sechs bis zehn Jahren zur nächstgelegenen Grundschule in Hessen dar? Zur Beant-

wortung dieser Frage wurde der Experimentaldatensatz des Mikrozensus Hessens um zusätzliche Merkmale erweitert. Das Besondere an diesen Merkmalen ist, dass sie nicht im Zuge der Befragung der Mikrozensus-Haushalte erhoben werden, sondern über den räumlichen Bezug – die Rasterzellen-ID – dem Haushaltsdatensatz nachträglich zugespielt werden. Aus Sicht der Statistik ist dies eine belastungsarme Methode zur Generierung zusätzlicher Merkmale. Die ermittelten Merkmale können schließlich, wie alle Merkmale des Mikrozensus, analysiert werden.

↘ Was ist der Mikrozensus?

Der Mikrozensus ist eine repräsentative Haushaltsbefragung der amtlichen Statistik in Deutschland. Seit 1957 wird im Rahmen des Mikrozensus jährlich 1% der Haushalte stellvertretend für die gesamte Bevölkerung zu ihren Lebensbedingungen befragt. In Hessen sind dies derzeit etwa 60 000 Personen in rund 30 000 Haushalten, die gleichmäßig über die Kalenderwochen des Jahres verteilt befragt werden. Bereits seit 1968 enthält das Frageprogramm des Mikrozensus auch Fragen der Arbeitskräfteerhebung der Europäischen Union. Um ein repräsentatives Bild der Lebenssituation der Gesamtbevölkerung zu erhalten, besteht für den Mikrozensus Auskunftspflicht.

3.2 Geokodierung

In einem ersten Schritt erfolgte die Geokodierung der hessischen Grundschulstandorte¹, das heißt die vorhandenen Anschriften wurden in Koordinatenpaare transformiert und die Schulen damit punktgenau verortet.

Die hessischen Mikrozensus-Haushalte wurden ebenfalls geokodiert. Der Datensatz jedes Mikrozensushaushalts wurde um Koordinaten erweitert. Ausgehend von den Koordinaten erfolgte anschließend die Datensatzergänzung um die Rasterzellen-ID, welche die Rasterzelle definiert, in der das Koordinatenpaar liegt. Die

¹ Die Anschriften der Grundschulen stammen aus dem öffentlich zugänglichen Hessischen Schulverzeichnis und basieren auf dem Hessischen Schulinformationssystem. Das Schulverzeichnis enthält nicht immer den tatsächlichen Standort der Grundschulen, weil das Verzeichnis die postalische Erreichbarkeit sicherstellen soll. In Stichproben wurden die tatsächlichen Grundschulstandorte über die betreffenden Schulwebseiten recherchiert. Die Nebenstandorte, die nicht im Schulverzeichnis enthalten sind, wurden vom Hessischen Kultusministerium aus dem Hessischen Dienststellenverzeichnis bereitgestellt.

Rasterzellen-ID wird gespeichert, die Koordinaten nicht (Scholze, 2015, hier: Seite 211).

3.3 Berechnung von Entfernungen

Um die Entfernungen zwischen den Haushalten und der nächstgelegenen Grundschule berechnen zu können, wurden die Wohnstandorte der Haushalte operationalisiert. Es wurden nicht die Entfernungen zwischen den Wohnstandorten und den nächstgelegenen Grundschulen gemessen, sondern die Entfernungen zwischen der Grundschule und der Rasterzelle, in der sich der Haushalt befindet. Das Operationalisieren auf Rasterebene dient der späteren Verknüpfung der Mikrozensus-Daten mit den berechneten Entfernungen.

Für die Distanzmessung wurden mithilfe eines Routingdienstes Erreichbarkeitszonen berechnet. Dabei wurde ein INSPIRE²-konformes 100-m-Raster für Hessen um Erreichbarkeitsinformationen angereichert. Ausgehend von den Schulstandorten wurden Erreichbarkeitszonen definiert. Erreichbarkeitszonen sind Gebiete, die von einem Startpunkt aus innerhalb eines bestimmten Zeitraums oder innerhalb einer bestimmten Wegstrecke (nicht Luftlinie!) erreichbar sind. Startpunkte waren die einzelnen Grundschulen. Ausgehend von diesen wurden Erreichbarkeitszonen (zum Beispiel unter 300 Meter, 300 bis unter 500 Meter und weitere Zonen) festgelegt. Die erzeugten Erreichbarkeitszonen und die 100-m-Rasterdatei für Hessen wurden anschließend zur weiteren Verarbeitung in ein Datenbanksystem importiert. Unter Verwendung räumlicher Operationen wurden auf Datenbankebene die Rasterinformationen mit den errechneten Erreichbarkeitszonen verschnitten. Je nachdem, in welche Erreichbarkeitszone dabei der Rastermittelpunkt fiel, erhielten die Rasterzellen den entsprechenden Wert (zum Beispiel 300 bis unter 500 Meter). Dieses Verfahren wurde flächendeckend durchgeführt. Abschließend verfügte jede Rasterzelle Hessens über Distanzwerte zur nächstgelegenen Grundschule. Jede Rasterzelle hatte zudem einen eindeutigen Identifikator, die Rasterzellen-ID.

² www.geoportal.de

Zur Berechnung der Erreichbarkeitszonen für die hessischen Grundschulen wurde der Routingdienst³ des Bundesamtes für Kartographie und Geodäsie eingesetzt. Dieser Dienst basiert auf dem OpenRouteService des „Heidelberg Institute for Geoinformation Technology“ (HeiGIT) der Universität Heidelberg. Als Datengrundlage des Routingdienstes wird die frei verfügbare Datenbasis „OpenStreetMap“ verwendet.⁴

Neben der klassischen Berechnung von Pkw-, Fahrrad- und Fußgängerrouen bietet der Dienst auch die Möglichkeit, Geh-/Fahrzeiten-Intervalle zu berechnen. Es ist also möglich, die Distanz in Strecke, wie auch in Dauer zu messen. Die Berechnung erfolgt stets entlang bestehender Verkehrswege. Nicht berücksichtigt wurden in der vorliegenden Analyse Ampeln, Baustellen, aktuelle Verkehrslagen (Echtzeitdaten) und zu erwartende Verkehrslagen (Tageszeit und Wochentag).

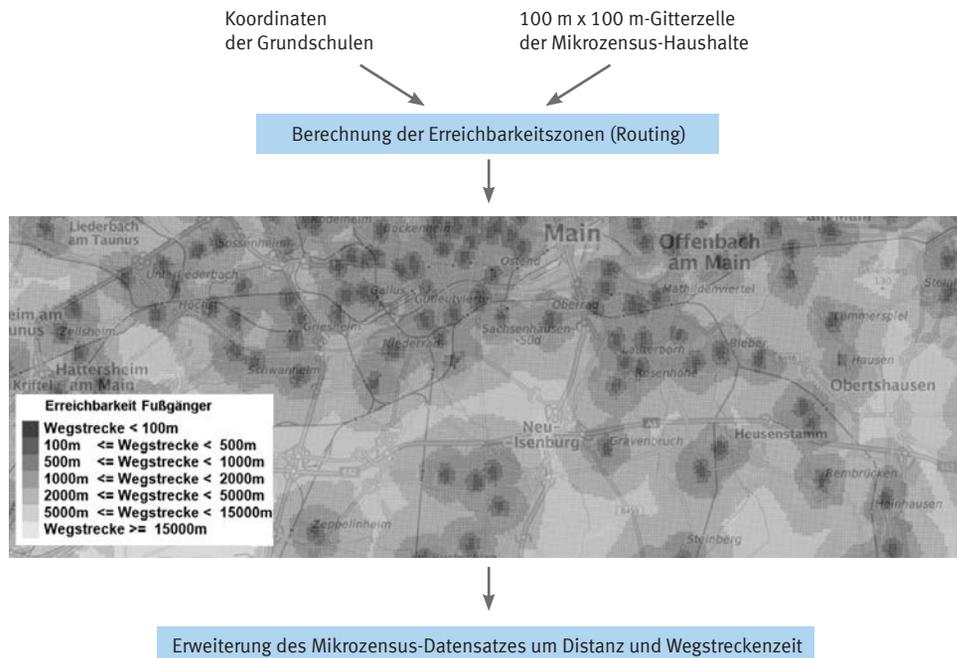
3.4 Zusammenführen der Informationen

Nach der Verschneidung der Rasterinformationen mit den Erreichbarkeitszonen liegt nun ein 100-m-Raster für Hessen vor, welches für alle Rasterzellen Informationen über die Distanz zur nächstgelegenen Grundschule enthält. Die einzelnen Rasterzellen verfügen über einen eindeutigen Identifikator, die Rasterzellen-ID. Über die Rasterzellen-ID ist eine Verknüpfung der Mikrozensus-Haushalte und der Distanzinformationen möglich. Die einzelnen Mikrozensusdatensätze wurden somit um neue Informationen bezüglich der Distanzen und Wegstreckenzeiten zur nächstgelegenen Grundschule ergänzt. [↗ Grafik 2](#)

- ³ Im Routingdienst sind verschiedene Geschwindigkeitsprofile für Pkw, Fahrräder und Fußgänger hinterlegt. Bei der Berechnung der Routen und Erreichbarkeitszonen zum Beispiel für Pkw und Fahrräder berücksichtigt der Routingdienst zudem die Art der Straße (Autobahn, Landstraße, Innenstadt und weitere Straßenführungen), die Geschwindigkeitsbegrenzung (im Normalfall durch Straßenschilder) und die Oberflächenbeschaffenheit (Asphalt, Schotter, Sand und weitere Straßenbeläge) der Straße. Die tatsächlich gewählte Geschwindigkeit für einen Straßenabschnitt ergibt sich dabei aus dem niedrigsten der drei Werte. Die Fußgänger-Geschwindigkeit ist im Normalfall unabhängig von der Beschaffenheit auf 5 km/h festgelegt. Dies bedeutet allerdings auch, dass für die Gesamtzeit des Wegs (das heißt die auf der jeweiligen Straße verbrachte Zeit) immer angenommen wird, dass die ausgewählte Geschwindigkeit durchgehend auf dem Streckenabschnitt eingehalten und durchgehalten wird.
- ⁴ Grund dafür ist, dass aktuell keine routingfähigen amtlichen Datengrundlagen existieren, mit denen ähnliche Berechnungen möglich sind.

Grafik 2

Erweiterung des Mikrozensus-Datensatzes: Fußläufige Erreichbarkeit der nächsten Grundschulen



Quelle: Visualisierung auf Basis der Ergebnisse des Routingdienstes des Bundesamtes für Kartographie und Geodäsie

2019 - 01 - 0519

3.5 Ergebnisse

Auf Basis des erstellten hessischen Projektdatensatzes für den Mikrozensus 2017 konnten exemplarisch Ergebnisse generiert werden. Dabei handelt es sich nicht um „amtliche“ Ergebnisse, sondern um experimentelle Daten. Zum einen konnten für das Projekt nur vorläufige Daten des Mikrozensus 2017 genutzt werden, zum anderen muss der Prozess der Qualitätssicherung noch weiter vorangetrieben werden. So müsste beispielsweise sichergestellt werden, dass die in der Analyse verwendeten Daten – Mikrozensus, Schulverzeichnis, Routinginformationen – das gleiche Bezugsjahr haben.

Nach der modellhaften Analyse wäre für rund 90% der hessischen Familien mit mindestens einem Kind im grundschulpflichtigen Alter von sechs bis zehn Jahren die nächstgelegene Grundschule höchstens zwei Kilometer entfernt. Gemessen wird dabei nicht die Luftlinie, sondern die über eine Straße zurückzulegende Entfernung. Weitere rund 10% der Familien mit mindestens einem Kind im grundschulpflichtigen Alter müssen maxi-

mal eine Strecke von fünf Kilometern zurücklegen, um eine Grundschule zu erreichen. Der Anteil der Familien mit einer Entfernung von mehr als fünf Kilometern ist so gering, dass er im statistischen Sinn nicht als valide angesehen werden kann.

Die Anteilswerte variieren zwischen eher städtisch und eher ländlich geprägten Regionen.⁵ In städtischen Gemeinden haben nahezu alle Familien mit Kindern im grundschulpflichtigen Alter (99%) eine solche Schule im Umkreis von zwei Kilometern. Dagegen trifft dies in ländlichen Gemeinden auf vergleichsweise wenige Familien zu, genauer auf einen Anteil von 55% und damit nur knapp über die Hälfte der Familien.

Bei der Betrachtung der berechneten Fahrzeit zur nächstgelegenen Grundschule zeigt sich, dass rund 97% der Familien mit einem Kind im grundschulpflichtigen Alter maximal fünf Minuten für die Fahrt zur nächsten Grundschule benötigen. In städtischen Regionen brauchen

5 Zum Verstädterungsgrad (Degree of urbanisation – DEURBA) siehe <https://ec.europa.eu/eurostat/de/web/degree-of-urbanisation/background>

alle Familien der betrachteten Gruppe (100%) maximal fünf Minuten Fahrzeit zur nächstgelegenen Grundschule, in ländlichen Regionen sind es mit 90% etwas weniger (Statistisches Bundesamt, 2018, hier: Seite 12 f.).

Für die Interpretation der Ergebnisse ist allerdings zu beachten, dass die Berechnung der Fahrzeiten auf einer modellbedingten Annahme basiert: Die gesamte Fahrt wird, vom ersten bis zum letzten Meter, mit der jeweils maximal erlaubten Höchstgeschwindigkeit zurückgelegt. Die Betrachtung der Fahrzeit zur nächstgelegenen Grundschule für Familien mit grundschulpflichtigen Kindern für städtische und ländliche Regionen hält einem Vergleich mit empirischen Erfahrungswerten nicht stand. Insbesondere im Stadtverkehr sind mitunter auch kurze Strecken zur entsprechenden Tageszeit kaum in dieser Fahrzeit zu schaffen. Dies spiegelt sich in der Auswertung der Pendlerdaten des Mikrozensus wider: Zwar legen Pendlerinnen und Pendler in städtischen Regionen häufiger kürzere Distanzen zur Arbeitsstätte zurück als in ländlichen Regionen. Jedoch ist dort der Anteil derjenigen, die eine kurze Fahrzeit von weniger als zehn Minuten zur Arbeit benötigen, geringer als auf dem Land (Dettmer/Wolf, 2018, hier: Seite 8).

4

Fazit und Ausblick

Die Georeferenzierung bestehender Datensätze ermöglicht es, räumlich generierte Merkmale zu ergänzen und somit neue Zusammenhänge zu analysieren – und zwar unabhängig davon, ob eine Vollerhebung oder eine Stichprobe zugrunde liegt. Sie kommt ohne zusätzliche Erhebungen und ohne zusätzliche Belastung der Befragten und des Haushalts aus. Darüber hinaus kann die Auswertungstiefe bestehender Erhebungen in der räumlichen Tiefe und Flexibilität der analysierten Regionaleinheiten gesteigert werden.

Das Potenzial räumlicher Analysen sollte weiter erforscht und genutzt werden, um zusätzliche statistische Informationen zu gewinnen und möglicherweise klassische Erhebungsmethoden durch solche Analysen teilweise zu ergänzen oder zu ersetzen. Zur Abschöpfung des Analysepotenzials geokodierter Statistiken sind sehr gute Kenntnisse der Fachstatistiken sowie der Verfahren der Geodatenverarbeitung und der zusätzlich verfügbaren

Geobasisdaten notwendig. In intensiven Kooperationen der Expertinnen und Experten der verschiedenen Bereiche können ein gegenseitiges Verständnis für jeweils anders gelagerte Perspektiven entwickelt und Analysen erarbeitet werden. Langfristig sollten bei entsprechendem Bedarf die exemplarisch durchgeführten räumlichen Analysen und Prozesse standardisiert und in die laufende Statistikproduktion integriert werden.

Die in diesem Artikel vorgestellte Analyse basiert auf einem Stadt-Land-Vergleich. Noch interessanter wären Aussagen auf lokaler Ebene. Jedoch reichen der Stichprobenumfang des Mikrozensus, die relativ geringe Anzahl der Befragungen (siehe Exkurs „Was ist der Mikrozensus?“) und die noch geringere Anzahl an Haushalten mit Kindern im Grundschulalter nicht aus, um valide Ergebnisse in tieferer räumlicher Gliederung auszuweisen. Für Auswertungen auf noch tieferer regionaler Ebene bräuchte es bei der hier behandelten Fragestellung neben den vorhandenen Anschriften der Grundschulen einen größeren Stichprobenumfang. Die benötigten Informationen, in diesem Fall Anschriften, finden sich in Verwaltungsregistern, zum Beispiel im Einwohnermelderegister. Um diese jedoch in größerem Umfang verwenden zu dürfen, müssen gesetzliche Grundlagen geschaffen werden.

Im Rahmen eines Routings können „tatsächliche“ Aufwände berechnet werden, die für bestimmte Distanzen benötigt werden. Distanzen können in der Einheit Zeit ausgewiesen werden und somit ein realistischeres Bild der Erreichbarkeiten abbilden als bei einer Betrachtung der Wegstrecke oder gar der Luftstreckendistanz. Verkehrsaufkommen oder naturräumliche Gegebenheiten werden zudem besser abgebildet. Das Routing kann auch unterschiedlichste Verkehrsmittel (Pkw, öffentlicher Personennahverkehr, zu Fuß, Fahrrad, ...) sowie für Fußgänger nachteilige Straßenbegebenheiten (wie steile Straßen) in die Analysen einbeziehen, um die in der Realität vorhandenen Mobilitätsoptionen widerzuspiegeln.¹⁶

Der Einsatz von Routingdiensten könnte im Verbund der Statistischen Ämter des Bundes und der Länder künftig auf viele weitere Fragestellungen ausgeweitet werden. Dazu zählen Informationen, wie gut andere Infrastruktureinrichtungen, beispielsweise weiterführende Schu-

⁶ Metadatenbeschreibung für den Indikator 11.2.1 der Sustainable Development Goals (Nachhaltigkeitsziele der Vereinten Nationen). [Zugriff am 9. Juli 2019]. Verfügbar unter: <https://unstats.un.org/>

len, Krankenhäuser⁷, öffentlicher Nahverkehr, oder die Versorgung mit Infrastruktureinrichtungen des täglichen Bedarfs durch Supermarkt, Friseur, Bäcker, Banken und so weiter, erreichbar sind. Solche Angaben ermöglichen es, Aussagen über Versorgungslagen zu treffen (Bundesamt für Statistik, 2018). Anschriften und Koordinaten dieser Einrichtungen sind bereits heute im statistischen Unternehmensregister vorhanden. Auch die Erreichbarkeit von Freizeitinfrastruktur, wie Seen, Parks und Wälder, könnte berechnet werden.

Des Weiteren sollten auch die vorhandenen, bei der vorliegenden Untersuchung eingesetzten Instrumente für die Erstellung valider statistischer Informationen weiterentwickelt werden. So ist eine Weiterentwicklung des „OpenRouteService“ des Bundesamtes für Kartographie und Geodäsie erforderlich. Die aufgezeigten Vereinfachungen führen bisher zu nicht realistischen Werten und damit zu gewissen Verzerrungen in den angegebenen Fahrzeiten.

Außerdem bedarf es einer noch exakteren Recherche der Grundschulstandorte (siehe Fußnote 1). Denn eine hohe Ergebnisqualität ist nur sicherzustellen, wenn alle existierenden Schulen oder sonstigen Infrastruktureinrichtungen und ihre genaue Verortung berücksichtigt werden.

Die genannten Punkte verdeutlichen, dass neben der Etablierung der rein technischen Prozesse zur Geokodierung und Verschneidung der Datenbestände zusätzliche Aspekte zur Einschätzung der Datenqualität weiter untersucht werden müssen. 

⁷ <https://krankenhausatlas.statistikportal.de/>

LITERATURVERZEICHNIS

Bundesamt für Statistik (Schweiz). *Dienstleistungen für die Bevölkerung und regionale Disparitäten. Distanzen zu den Alltagsdienstleistungen im Jahr 2015*. In: BFS Aktuell. November 2018. Seite 1 ff.

Dettmer, Bianka/Wolf, Isabel. *Mobilität der hessischen Bevölkerung*. In: Staat und Wirtschaft in Hessen. Ausgabe 2/2018, Seite 3 ff.

Scholze, Sarah. *Statistische Ergebnisse „im Quadrat“. Geografische Raster am Beispiel des Zensus 2011*. In: Staat und Wirtschaft in Hessen. Ausgabe 5/2015, Seite 211 ff.

Statistische Ämter des Bundes und der Länder. *Digitale Agenda des Statistischen Verbunds. Version 1.0*. (Internes Dokument). 2018.

Statistisches Bundesamt. *Analyse georeferenzierter Daten. Abschlussbericht zu den Arbeiten des Projektteams*. (Internes Dokument). 2018.

Van Halderen, Gemma/Minchin, Stuart/Brady, Martin/Scott, Greg. *Integrating statistical and geospatial information, cultures and professions: International developments and Australian experience*. In: Statistical Journal of the IAOS. Jahrgang 32. 2016. Seite 457 ff. DOI 10.3233/SJI-161018.

United Nations Expert Group on the Integration of Statistical and Geospatial Information. *Global Statistical Geospatial Framework: Linking Statistics and Place. Current status and plans for development, July 2018*. Sitzungsunterlage zur 8. Plenarsitzung der Experten des Global Geospatial Information Management zum Global-Statistical-Geospatial-Framework. New York 2018. [Zugriff am 10. Juli 2019]. Verfügbar unter: <http://ggim.un.org>

FORTSCHRITTE BEIM EINSATZ VON WEB SCRAPING IN DER AMTLICHEN VERBRAUCHERPREISSTATISTIK – EIN WERKSTATTBERICHT

Christian Bladow, Daniel Seeger

↳ **Schlüsselwörter:** Verbraucherpreisindex – Web Scraping – Digitalisierung – dynamische Preissetzung

ZUSAMMENFASSUNG

Für den deutschen Verbraucherpreisindex und den Harmonisierten Verbraucherpreisindex werden derzeit monatlich für etwa 10 000 Produkte Preise im Internet erhoben. Im Regelfall erfolgt dies manuell durch Preiserheberinnen und Preiserheber, und zwar zu einem festgelegten Zeitpunkt. Im Statistischen Bundesamt wird intensiv daran gearbeitet, diese manuellen Preiserhebungen bis Ende 2020 mit der Hilfe von Web Scraping zu automatisieren. Auch werden neue Methoden entwickelt, um große Datenmengen in die Berechnung von Preisindizes zu integrieren. Der vorliegende Artikel stellt das in der Entwicklung befindliche generische Programm für die Nutzung von Web Scraping vor und gibt darüber hinaus einen Überblick über den Forschungsstand in der Methodenentwicklung.

↳ **Keywords:** consumer price index – web scraping – digitalisation – dynamic pricing

ABSTRACT

At present, the prices of roughly 10,000 products are collected on the internet each month for the German Consumer Price Index and the Harmonised Index of Consumer Prices. Usually, this is done manually by price collectors, and at a given point of time. At the Federal Statistical Office, intensive work is underway to automate the manual price collection by means of web scraping by the end of 2020. Also, new methods are being developed to include large data volumes in the calculation of price indices. This article presents the generic program that is being developed for the use of web scraping and gives an overview of the state of research in the field of methods development.



Christian Bladow

hat International Economics an der Universität Göttingen studiert und ist Referent im Referat „Zwischenörtliche Preisvergleiche, Parteienindex“ des Statistischen Bundesamtes. Schwerpunkte seiner Arbeit sind die Automatisierung von Preiserhebungen im Internet sowie die Entwicklung von Methoden für die Integration von und den Umgang mit großen Datenmengen.



Daniel Seeger

ist Diplom-Volkswirt und leitet das Referat „Verbraucherpreise“ im Statistischen Bundesamt. Er ist insbesondere zuständig für die Koordinierung der Zusammenarbeit mit den Statistischen Ämtern der Länder und arbeitet auf europäischer Ebene an der Weiterentwicklung der Methoden des Harmonisierten Verbraucherpreisindex mit.

1

Einleitung

Der Internethandel gewinnt stetig an Bedeutung. Dementsprechend erfasst das Statistische Bundesamt für die Verbraucherpreisstatistik zunehmend Preise der im Internet gehandelten Waren und Dienstleistungen. Derzeit werden für die Verbraucherpreisstatistik jeden Monat Preise für etwa 10 000 Produkte im Internet erhoben, überwiegend manuell. Der Anteil der Produkte, für die Preise im Internet erhoben werden, hat sich im Verbraucherpreisindex (VPI) beziehungsweise im harmonisierten Verbraucherpreisindex (HVPI) seit dem Basisjahr 2005 bis zum Basisjahr 2015 von 5 % auf mehr als 10 % erhöht. Diese Aussage gilt zumindest bei den Gütern des Warenkorb, für die eine Einteilung der Konsumausgaben der privaten Haushalte in verschiedene Geschäftstypen¹ (Sandhop, 2012) möglich ist. Der steigende Anteil resultiert auch daraus, dass große Einzelhandelsunternehmen neben ihrem Verkauf in stationären Geschäften ihre Produkte verstärkt auch in Onlineshops anbieten. Zum Teil reduzieren sie ihr Sortiment im stationären Handel sogar zugunsten des Internethandels. Der steigende Anteil der Waren und Dienstleistungen, die im Internet gehandelt werden, hat zur Folge, dass die Stichprobe der Preiserhebung im Onlinehandel künftig weiter ausgeweitet werden muss.

Parallel wird in der Preisstatistik daran gearbeitet, die Frequenz der Preiserhebungen zu optimieren. Studien über das Preissetzungsverhalten von Onlinehändlern haben gezeigt, dass sich Preise im Onlinehandel häufiger ändern als im stationären Handel (Blaudow/Burg, 2018). Eine höhere Frequenz der Preiserhebungen ist nur durch einen verstärkten Einsatz automatisierter Verfahren wie Web Scraping zu erreichen. Zusätzlich werden neue Methoden entwickelt, um die zu erwartenden Datenmengen zu validieren und in die monatliche Indexberechnung des VPI und des HVPI zu integrieren.

Der vorliegende Beitrag gibt einen Überblick über die anstehenden Veränderungen in der Internetpreiserhebung, sowohl aus technischer als auch aus methodischer Sicht.

1 Hierbei muss festgehalten werden, dass sich die Geschäftstypengewichtung nur auf etwa ein Drittel aller Güter im Warenkorb bezieht.

2

Der Einsatz von Web Scraping in der Verbraucherpreisstatistik

Seit dem Jahr 2012 werden im Statistischen Bundesamt die Möglichkeiten automatisierter Preiserhebungen im Internet mithilfe von Web Scraping untersucht. Finanziert wurden und werden diese Arbeiten weitgehend durch das Statistische Amt der Europäischen Union (Eurostat). Dabei wurde von Beginn an ein imitierender Ansatz gewählt, bei dem durch Web Scraping einzelne, manuelle Preiserhebungen im Internet automatisiert werden. Das heißt, zu bestimmten Zeitpunkten werden automatisiert Preise für zuvor definierte Produkte extrahiert, die ansonsten manuell erhoben wurden. Die Internetseite, auf der das jeweilige Produkt angeboten wird, kann entweder mit der entsprechenden URL² direkt aufgerufen werden. Oder es wird anhand von Informationen, die in einer Eingangsdatenbank liegen oder im Programm errechnet werden, Schritt für Schritt zur Internetseite gesteuert. Das Programm erkennt anhand der festgelegten Definition das Produkt und extrahiert den gesuchten Preis. Somit wird die manuelle Arbeit der Preiserheberinnen und Preiserheber imitiert, ist jedoch in beliebiger Frequenz wiederholbar. Im Gegensatz dazu könnten auch die Preise aller verfügbaren Produkte beispielsweise eines Onlinehändlers ohne vorausgehende Produktauswahl³ abgegriffen werden. Dieser Ansatz wird vom Statistischen Bundesamt bislang nicht verfolgt.

Derzeit wird Web Scraping in der Preisstatistik isoliert vom hausinternen Netzwerk des Statistischen Bundesamtes mit projekteigener IT-Ausstattung betrieben. Das eigentliche Programm, welches die Automatisierungsprozesse steuert, ist in der Programmiersprache Java geschrieben. Es verwendet SQL⁴-Datenbanken für Eingangs- und Ausgangsdaten sowie entweder das Scrapingtool iMacros oder Selenium. Die Programmierung kann ebenso wie die Programm- und Datenpflege nur von wenigen Personen mit speziellen IT-Kenntnissen durchgeführt werden.⁵

2 Uniform Resource Locator – einheitlicher Ressourcenzeiger. Die URL ist ein Standard, mit dem bestimmte Inhalte aufgerufen werden; häufig wird darunter die Adresse einer Website verstanden.

3 Sogenanntes bulk scraping.

4 Structured Query Language.

5 Für zusätzliche Erläuterungen siehe Blaudow (2018).

Aktuell werden bereits für mehrere Güterbereiche Preise für die Verbraucherpreisstatistik erfolgreich mittels Web Scraping erhoben, beispielsweise für Mietwagen, für zahlreiche Arten von Medikamenten oder auch für Fernbus- und Bahnreisen.

Über die Preiserhebungen für den VPI/HVPI hinaus wird Web Scraping auch in anderen Bereichen der Preisstatistik erfolgreich eingesetzt:

- › Für die Kaufkraftparitäten werden Preise für Hotelübernachtungen und Flugreisen erhoben.
- › Für den Preisindex des selbstgenutzten Wohneigentums werden Wohngebäudeversicherungsstarife erhoben.
- › Für den Erzeugerpreisindex werden Gas- und Strompreise von mehreren Energiebörsen automatisiert aus dem Internet gewonnen.

Je nach Statistik und Volatilität der Preise werden die automatisierten Preiserhebungen täglich, wöchentlich oder monatlich durchgeführt.

3

Die Herausforderung durch dynamische Preissetzung im Onlinehandel

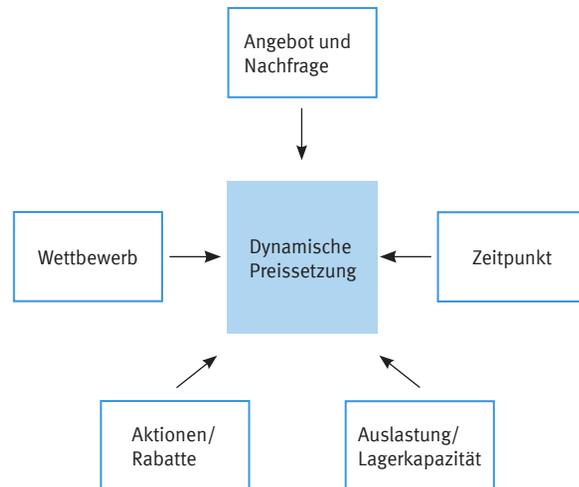
Im Internethandel ist es mit wenig Aufwand möglich, Preise anhand von Algorithmen in sehr kurzen Zeitabständen automatisiert und annähernd unbegrenzt zu ändern. Dieses Phänomen wird allgemein als dynamische Preissetzung bezeichnet. Die hierfür verwendeten Algorithmen werden durch Faktoren wie Angebot und Nachfrage, Zeitpunkt des Kaufs und Rabattaktionen beeinflusst.¹⁶ [↪ Grafik 1](#)

In einem Markt mit einer statischen Preisgestaltung (siehe [↪ Grafik 2](#)) bieten die Produzenten ihre Produkte zu einem Preis an, der für alle Kundinnen und Kunden gleich ist und für einen absehbaren Zeitabschnitt (beispielsweise über einen Monat) gleich bleibt. Ein Produzent kennt nicht die individuelle Zahlungsbereitschaft

¹⁶ In der Studie von Blaudow/Burg (2018) wurde beispielweise festgestellt, dass ein Großteil der Preisänderungen bei größeren Online-shops während der ersten Tageshälfte auftritt.

Grafik 1

Auswahl an Einflussfaktoren für dynamische Preissetzung

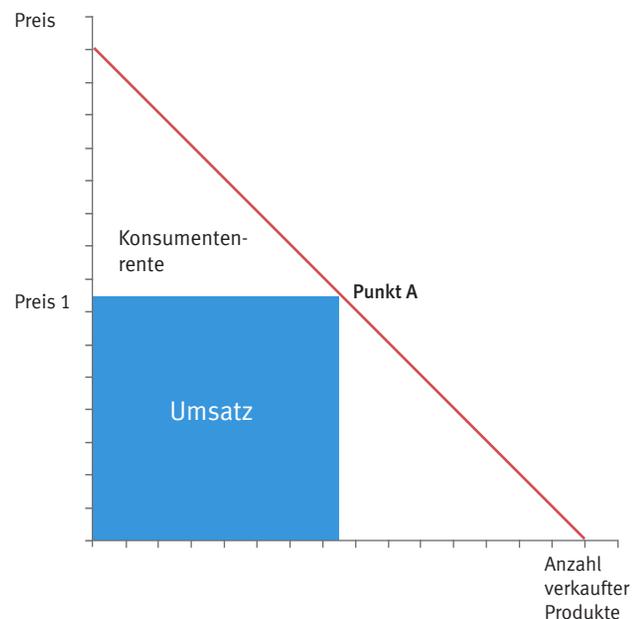


2019 - 01 - 0521

der Kundschaft und wird einen Preis 1 setzen, der möglichst viele Kundinnen und Kunden zum Kauf seines Produkts veranlasst. Die Kundschaft kann auf diese Weise das Produkt und dessen Preis mit Produkten anderer

Grafik 2

Statische Preisgestaltung



2019 - 01 - 0522

Produzenten vergleichen, um ihre Konsumentenrente¹⁷ zu maximieren.

Liegt die individuelle Zahlungsbereitschaft einer Kundin oder eines Kunden oberhalb des gesetzten Preises, ist die Konsumentenrente beim Kauf des Produkts positiv. Liegt der Preis hingegen oberhalb der individuellen Zahlungsbereitschaft, würde sich die Kundin oder der Kunde gegen einen Kauf entscheiden und möglicherweise ein Substitut bei einem anderen Produzenten zu einem geringeren Preis erwerben.

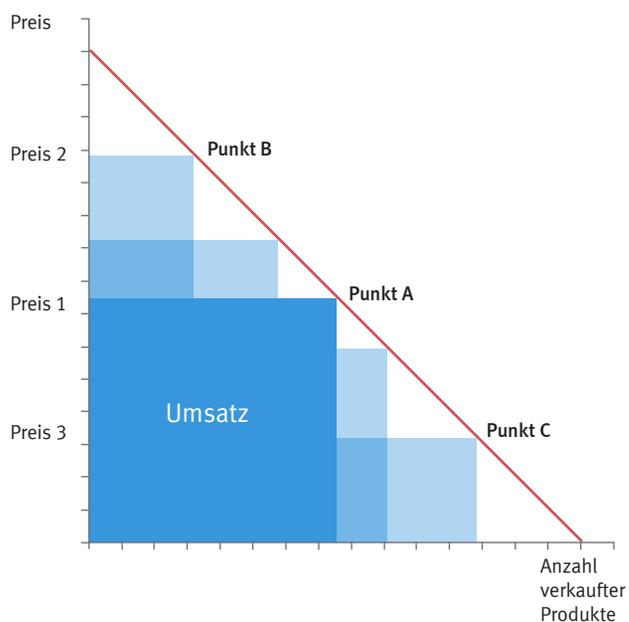
Mit einer dynamischen Preissetzungsstrategie kann gegenüber einer statischen Preissetzung der Gewinn des Verkäufers ausgeweitet werden: Dazu müsste es dem Verkäufer gelingen, die Preise so zu setzen, dass sie idealerweise genau die Zahlungsbereitschaft des potenziellen Käufers treffen. Kundinnen und Kunden mit hoher Zahlungsbereitschaft würden einen höheren Preis zahlen (Preis 2 in [Grafik 3](#)) entsprechend ihrer Zahlungsbereitschaft und die Konsumentenrente ten-

diert gegen null. Kundinnen und Kunden mit einer Zahlungsbereitschaft unter dem statischen Preis würde ein entsprechend niedrigerer Preis (Preis 3) angeboten. Sie würden zum niedrigeren Preis kaufen und nicht auf ein Substitut ausweichen. Der Absatz würde steigen, der Gewinn des Verkäufers optimiert und die Konsumentenrente abgeschöpft.

Im Onlinehandel ermöglichen die verfügbaren Algorithmen sehr schnelle Preisanpassungen. Ein Wechsel zwischen verschiedenen Produktpreisen (Preis 1, Preis 2 und Preis 3) ist ohne großen Aufwand durchzuführen und jederzeit wiederholbar, wodurch ein Produzent seine Absatzstrategie stets den Marktgegebenheiten anpassen kann.

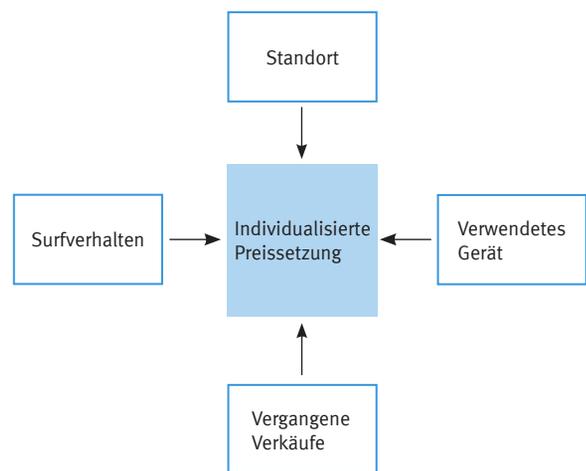
Das Phänomen der dynamischen Preissetzung ist zu trennen vom Phänomen der individualisierten Preissetzung. Bei den Algorithmen für die individualisierte Preissetzung sind personenabhängige Eigenschaften die entscheidenden Kriterien, die die Preissetzung beeinflussen. [Grafik 4](#)

Grafik 3
Dynamische Preisgestaltung



2019 - 01 - 0523

Grafik 4
Auswahl an Einflussfaktoren für individualisierte Preissetzung



2019 - 01 - 0524

⁷ Als Konsumentenrente bezeichnet man in der Wohlfahrtstheorie die Differenz zwischen der individuellen Zahlungsbereitschaft einer Käuferin oder eines Käufers für ein Produkt und dem tatsächlich gezahlten Preis beim Kauf [siehe beispielsweise Varian (2010), hier: Seite 264].

Sowohl das Phänomen der dynamischen als auch das der individualisierten Preissetzung im Internet stellt eine Herausforderung für die amtliche Preisstatistik dar. Stark volatile Preise machen eine repräsentative Preismessung schwieriger. Während die individualisierte Preissetzung konzeptionell bei der Inflationsmessung noch

nicht berücksichtigt werden soll, ist eine dynamische Preissetzung in der durchschnittlichen Preisentwicklung korrekt zu erfassen. Deshalb wird die Preisvolatilität bei den für die Verbraucherpreisstatistik relevanten Produkten kontinuierlich untersucht. Derzeit wird in einer Studie mit rund 9 000 Produkten das Preissetzungsverhalten von etwa 250 Onlinehändlern beobachtet und analysiert. Dazu werden stündlich Preise erhoben und ausgewertet. Ziel der Studie ist es, die Preisvolatilität der einzelnen Produkte festzustellen und bei Bedarf die Preiserhebung zu verbessern, um die Repräsentativität und die Qualität der monatlichen Preiserhebung des VPI/HVPI zu sichern oder sogar zu steigern.

4

Die Zukunft des Web Scrapings in der Preisstatistik: das generische Programm

Es ist vorgesehen, mit Web Scraping sukzessive die bisherigen manuellen Preiserhebungen im Internet abzulösen. Damit können in kurzer Zeit größere Datenmengen erhoben werden und die Datenerhebungen werden flexibler. So können beispielsweise Erhebungen an Wochenenden oder Nacherhebungen ohne großen Aufwand durchgeführt werden. Allerdings kann die Wartungsintensität hoch ausfallen, wenn eine Internetseite häufig verändert wird (technische Wartung) oder die zu erhebenden Produkte häufig wechseln und ersetzt werden müssen (fachliche Wartung).

Die fachliche Wartung erfordert unter anderem regelmäßige Kontrollen der Input- und Outputdatenbanken, da das Produktsortiment im Onlinehandel durch häufige Wechsel geprägt ist. Hierzu ist geschultes Personal in den Fachreferaten erforderlich, welches die automatisiert erhobenen Preisdaten auf deren Plausibilität hin überprüft und bei auftretenden Schwierigkeiten die notwendigen Gegenmaßnahmen einleitet (zum Beispiel Produkte manuell nacherheben). Die qualitativen Anforderungen an die Mitarbeiterinnen und Mitarbeiter steigen dadurch. Die Preiserhebungen im Internet, die bisher überwiegend einmal monatlich für jedes Produkt durchgeführt werden, erfolgen künftig automatisiert und in vielen Fällen in einer sehr viel höheren Frequenz. Die bestehende Stichprobe der etwa 10 000 Produkte

im Onlinehandel muss dabei nicht ausgeweitet werden, sondern entsprechend der Volatilität der Preise beziehungsweise der Preissetzungsstrategien der Onlinehändler werden mehr Preise je Produkt erfasst. Plausibilisierungsschritte, die bislang im Zuge der Preiserhebung erledigt wurden, müssen nachgelagert und sehr zeitnah im Erhebungsprozess für einen deutlich größeren Datensatz erfolgen.

Um die technische Wartung zu vereinfachen, wird seit Mitte des Jahres 2018 ein generisches Programm für Web Scraping in der Preisstatistik entwickelt. Mit dessen Hilfe sollen der Wartungsaufwand und die notwendigen IT-Kenntnisse für Web Scraping deutlich reduziert werden. Ziel ist, dass Web Scraping ohne tiefergehende IT-Kenntnisse von vielen Mitarbeiterinnen und Mitarbeitern eingesetzt und gesteuert werden kann. Ein technisches Grundverständnis über den Aufbau von Internetseiten oder den Umgang mit Datenbanken wird dennoch auch weiterhin unabdingbar sein. Spezielle Schulungen werden die Einführung des generischen Programms begleiten.

Das generische Programm wird als nutzerfreundliche Anwendung konzipiert: Die Benutzerinnen und Benutzer müssen keine Codes programmieren, sondern lediglich Abläufe anlegen, Variablen definieren und HTML-Positionen der Informationen auf Internetseiten in entsprechenden Masken einfügen. Eingangsdaten, die zur Identifizierung des zu erhebenden Produkts notwendig sind, können außerhalb der Anwendung gepflegt und aktualisiert werden und für die Erhebung in die Anwendung geladen werden. Die Anwendung wird darüber hinaus in der Lage sein, aus den definierten Variablen automatisch zusätzliche Variablen für die weitere Verwendung im Programm anzulegen. Die Ausgabedaten, also die Ergebnisse der automatischen Preiserhebungen, werden in einer Datenbank abgelegt und können über einen Browser evaluiert werden. Der gesamte Erhebungsprozess einschließlich der Navigation im Internet wird von einem herkömmlichen Internetbrowser übernommen, um Ladezeiten und Skriptfehler zu minimieren.

Grafik 5

Prozess der Datenerhebung mithilfe des generischen Programms



2019 - 01 - 0525

5

Prozess der Datenerhebung mithilfe des generischen Programms

Mit dem Einsatz des generischen Programms muss der Prozess der Datenerhebung im Vergleich zum bisherigen Vorgehen erweitert werden:

Zu Beginn sind die wesentlichen Informationen zum Produkt, wie Name, Artikelnummer und exakte Internetadresse (URL), in einer Tabelle zusammenzufassen. Hinzu kommen wichtige Informationen für die spätere Weiterverwendung, wie die (COICOP⁸-)Produktnummer, die Berichtsstellenummer und die Produktvariantennummer. Diese auch bisher schon erforderlichen Eingangsdaten werden in die Anwendung importiert und für die weitere Verwendung als Variablen angelegt. Daraufhin kann die Benutzerin oder der Benutzer den Ablauf der Preiserhebung definieren. So besteht die Möglichkeit, das Programm direkt zur Produktseite navigieren zu lassen oder schrittweise durch mehrere Seiten zu steuern (beispielsweise über die Auswahl mehrerer Filter). Schleifen und Bedingungen können ebenfalls angelegt werden. Mit Schleifen und Bedingungen kann die Datenerhebung präziser ablaufen und die Qualität der Ergebnisse gesteigert werden. In einem Zeitplan kann die Benutzerin oder der Benutzer festlegen, ob die Erhebung stündlich, täglich, wöchentlich oder jährlich wiederholt werden soll. Die Tage und genauen Uhrzeiten können ebenfalls eingestellt werden.

Unmittelbar nach einer Datenerhebung oder spätestens zum Ende eines gesamten Berichtszeitraums werden die erhobenen Daten von Preiserheberinnen und Preiserhebern validiert. Dieser Prozess ist in vier Schritte unterteilt, die im nachfolgenden Kapitel näher beschrieben

⁸ COICOP: Classification of Individual Consumption of Purpose – Klassifikation des individuellen Verbrauchs nach Verwendungszwecken.

werden. Im Anschluss an die Validierung werden die plausiblen Preise entweder direkt in das Berechnungsprogramm der Verbraucherpreisstatistik übernommen oder es müssen Preisindizes berechnet werden, welche als sogenannte Messzahlen in den weiteren Produktionsprozess gegeben werden. Damit wäre dann ein Erhebungsprozess abgeschlossen und kann wieder von vorne beginnen. [↪ Grafik 5](#)

6

Neue Validierungsprozesse in der Verbraucherpreisstatistik

Die Verwendung großer Datenmengen in der monatlichen Statistikproduktion setzt einen umfassenden und effizienten Validierungsprozess voraus. Insbesondere bei der Preiserhebung im Internet können unplausible Ergebnisse das Gesamtergebnis verfälschen. Aus diesem Grund müssen dem aktuellen Validierungsprozess in der Verbraucherpreisstatistik, der auf eine manuelle Preiserhebung ausgerichtet ist, weitere Validierungsprozesse vorgeschaltet werden.

6.1 Validierung durch das Scrapingprogramm

Die Validierung beginnt schon während der automatisierten Preiserhebung. Das generische Programm wird in der Lage sein, bestimmte Charakteristika eines Produkts zu prüfen, bevor der dazugehörige Preis erhoben wird. Solche Charakteristika können Artikelnummern der zu erhebenden Produkte oder – zum Beispiel beim Personentransport – die Verbindungsmerkmale sein. Dieses Vorgehen erhöht die Präzision der automatisierten Preiserhebung beträchtlich und reduziert die ungeeigneten Daten in der Ausgangsdatenbank.

6.2 Technische Validierung

Nach Abschluss der Preiserhebung erzeugt das generische Programm automatisch eine Datei mit einer Zusammenfassung des gerade abgeschlossenen Erhebungslaufs, in der Kennzahlen (zum Beispiel die Laufzeit der Erhebung, die Anzahl der erhobenen Preise, erkannte Fehler oder Warnungen) zusammengestellt werden. Darüber hinaus werden für zuvor definierte Fehlerarten Screenshots abgelegt. Diese Informationen ermöglichen es der oder dem für den Ablauf der Erhebung verantwortlichen Mitarbeiterin oder Mitarbeiter, die technische Wartung durchzuführen. Bei Bedarf kann ein vollständiger Ersatzlauf gestartet werden, um entstandene Lücken der Erhebung zu schließen.

6.3 Fachliche Datenvalidierung

Nach einem aus technischer Sicht erfolgreichen Erhebungslauf kann die Datenvalidierung durch Produktexpertinnen und Produktexperten beginnen. Vorübergehend nicht verfügbare Produkte sind im stationären und im Onlinehandel normal. Um im Nachgang der Erhebung beurteilen zu können, ob ein Produkt nur vorübergehend nicht verfügbar war oder dauerhaft aus dem Sortiment genommen wurde, werden Informationen zur Verfügbarkeit des Produkts in der Ausgangsdatenbank abgelegt. Sollte ein Produkt mehr als zwei Monate nicht verfügbar sein, wird das Produkt durch die zuständige Produktexpertin oder den zuständigen Produktexperten ersetzt. Auch bei fehlerhaft vorgegebenen Positionen der zu erhebenden Informationen auf den Produktseiten oder gänzlich fehlerhaften URLs sind die Datenbanken entsprechend zu pflegen und für die nächsten Erhebungen vorzubereiten. Die erhobenen Preisdaten werden anschließend auf Ausreißer und nicht plausible Werte hin überprüft. Nicht plausible Werte sind manuell aus dem Datensatz zu entfernen. Bei großem Datenverlust – hervorgerufen beispielsweise durch technische Schwierigkeiten – muss an dieser Stelle notfalls imputiert werden, sofern eine Wiederholung des Erhebungslaufs nicht möglich ist.

6.4 Weiterverwendung von Web-Scraping-Preisindizes

Die monatliche Datenaufbereitung und Datenvalidierung der Verbrauchpreisstatistik insgesamt wird einheitlich in einem gemeinsamen Berechnungsprogramm der Länder und des Bundes durchgeführt. Dieses sogenannte Verbundprogramm⁹ ist auf eine traditionelle Preiserhebung ausgerichtet, daher akzeptiert es für die monatliche Indexberechnung nur einen einzelnen Preis oder eine Messzahl für das jeweilige Produkt. Messzahlen sind außerhalb des Erhebungsprogramms zu berechnen und anschließend im Verbundprogramm zu erfassen.

Aktuell werden in der Preisstatistik neue Methoden getestet, wie aus großen Datenmengen [die zum Beispiel mittels Web Scraping oder auch mithilfe anderer moderner Datenquellen wie Scannerdaten (Bieg, 2019) gewonnen werden] verwertbare Preisindizes auf Produktebene für die monatliche Indexproduktion berechnet werden können.

Im Folgenden werden einige Mittelwerte erläutert und erste Ergebnisse dieser Untersuchungen vorgestellt.

7

Methodenentwicklung

Die automatisierten Preiserhebungen ermöglichen es, fast unendlich viele Preise aus dem Internet zu erfassen. Wie bereits erwähnt, ist dies insbesondere in solchen Produktbereichen beziehungsweise bei solchen Online-shops zwingend nötig, bei denen die Preise sehr volatil sind. Der Wandel von der traditionellen Indexberechnung mit monatlich einem Preis je Produkt hin zu mehreren Preisen je Produkt und Monat muss mit statistischen

Übersicht 1

Methoden zur automatisierten Preiserhebung

Methode	In Produktion	Experimentell
Modalwert	X	
Ungewichteter Durchschnitt	X	
Gewichteter Durchschnitt	X	
COICOP-Index		X

9 Für zusätzliche Erläuterungen siehe Burg/Seeger (2009).

Methoden begleitet werden. Drei einfache Methoden zur Durchschnittsberechnung werden bereits in der monatlichen Indexproduktion der Verbraucherpreisstatistik eingesetzt, eine Methode zur Aggregation großer Datenmengen wird auf ihre Anwendbarkeit geprüft. [↪ Übersicht 1](#)

7.1 Modalwert

Als Modalwert bezeichnet man den Wert, der am häufigsten in der Erhebung vorkommt. Diese Methode wird derzeit für Preiserhebungen verwendet, die im wöchentlichen Turnus durchgeführt werden. Bei monatlichen Statistiken werden also maximal vier Werte herangezogen, bei Quartalsstatistiken maximal zwölf Werte. Durch die Verwendung des Modalwerts einer Preiserhebung fließt ein tatsächlicher Preis in die Indexberechnung ein, die anderen Preise eines Produkts dienen der Plausibilisierung. Die Verwendung dieser Methode setzt voraus, dass die Preise nicht extrem volatil sind.

7.2 Ungewichteter Durchschnitt

Ein ungewichteter Durchschnittswert \bar{x} ist geeignet, wenn die Käufe eines beobachteten Produkts im Zeitverlauf gleich verteilt sind. Diese Methode wird verwendet, wenn weder der Wochentag noch die Tageszeit des Verkaufs gesondert gewichtet werden müssen. Es wird angenommen, dass sich die Preise nicht datumsabhängig verändern.

$$(1) \quad \bar{x} = \sum_{i=1}^n x_i$$

7.3 Gewichteter Durchschnitt

Die Verwendung eines gewichteten Durchschnitts \bar{x}_w ist geeignet, wenn zu den erhobenen Preisen sehr unterschiedlich häufig gekauft, aber gleichmäßig erhoben wurde. Beispielsweise ist anzunehmen, dass Reisende schnelle Zugverbindungen bevorzugen, dass Tickets in der 2. Klasse häufiger gekauft werden als Tickets in der 1. Klasse (zumindest zu bestimmten Tagen oder Zeiten) oder dass bestimmte Verbindungen (zum Beispiel Frankfurt – Köln) gefragter sind als andere. In der Berechnung

des Durchschnittspreises würde demnach eine Gewichtung w_j für mehr Repräsentativität sorgen.

$$(2) \quad \bar{x}_w = \sum_{i=1}^n w_i x_i$$

7.4 COICOP-Index

Neben der Anwendung dieser klassischen statistischen Methoden werden in der Preisstatistik weitere Ansätze zur Integration von großen Datenmengen untersucht. Ein Ansatz beispielsweise besteht darin, bei volatilen Preissetzungsverhalten dennoch einen repräsentativen Produktpreis für den Betrachtungszeitraum zu ermitteln, der möglichst alle (relevanten) Preisveränderungen erfasst. Mehrere statistische Ämter in der EU forschen an neuen Methoden zur Verwendung von großen Datenmengen im VPI/HVPI (Nygaard, 2015; van Loon/Roels, 2018; Griffioen/ten Bosch, 2016). Aktuell werden diese Methoden in der deutschen Preisstatistik evaluiert und weiterentwickelt beziehungsweise angepasst.

Das niederländische Statistikaamt verwendet beispielsweise einen sogenannten lowest COICOP-Index (Griffioen/ten Bosch, 2016). Dort wurde festgestellt, dass sich dieser besonders für Produktgruppen mit vielen Produktvarianten und gleichzeitig häufigen Produktwechseln eignet (zum Beispiel bei Bekleidung).

Ein analoger Ansatz wird in der deutschen Preisstatistik getestet: Der COICOP-Index wird aus den arithmetischen Mittelwerten der online erhobenen Preise für eine Produktgruppe berechnet, gewichtet mit den Anteilen eines Onlinehändlers für eine Produktgruppe. Als Produktgruppe wird die niedrigste COICOP-Produktklassifikation (in Deutschland die COICOP-10-Steller-Ebene) angesehen, weil dafür eine Gewichtung vorliegt.

Der COICOP-Index für die Produktgruppe g des Onlinehändlers r im Monat m wird auf folgende Weise berechnet:

$$(3) \quad CI_{G,m}^r = \sum_{g \in G} w_{g,m}^r I_{g,m}^r$$

Die Gewichtung einer Berichtsstelle wird durch w^r ausgedrückt und ergibt sich aus dem Anteil der Produkte eines Onlineshops an der Stichprobe des VPI/HVPI.

Die Stichprobe der für den VPI/HVPI relevanten Onlinehändler wird bisher im Zuge der fünfjährigen Revision anhand der Umsatzzahlen der Onlinehändler angepasst. Dazu werden auch Daten aus der Marktforschung verwendet. Die Gewichtung nach Onlinehändlern ist in solchen Produktbereichen sinnvoll, in denen bei vielen Onlinehändlern Preise erhoben werden und auch viele Produktvarianten erfasst werden (zum Beispiel Bekleidung). Außerdem ist sie sinnvoll bei Onlinehändlern, die zwar ein breites Angebot anbieten, aber nicht in allen Produktbereichen gleich führend sind (zum Beispiel Amazon).

Der Elementarindex $I_{g,m}^r$ stellt einen Elementarindex einer Produktgruppe g beim Onlinehändler r dar, beispielsweise Herrenjeans oder Damenfreizeitschuhe.¹⁰ Dieser Elementarindex wird mit Preisdaten berechnet, die mit der Hilfe von Web Scraping erhoben wurden.

$$(4) \quad I_{g,m}^r = \frac{P_{g,m}^r}{P_{g,m-1}^r}$$

Der Preis $P_{g,m}^r$ einer Produktgruppe g ergibt sich aus dem arithmetischen Mittel aller Preise, welche im Monat m beim Onlinehändler r erhoben wurden. Die Anzahl der verwendeten Preise ist besonders hoch, weil die Preise stündlich (h) erhoben werden. Es erfolgt keine Gewichtung nach Tageszeit oder Wochentag.¹¹

$$(5) \quad P_{g,m}^r = \frac{1}{n_{g,m}^r} \sum_{h \in m} \sum_{a \in g} p_{a,m,h}^r$$

Der Faktor p repräsentiert den Preis eines Produkts a , der auf der Internetseite eines Onlinehändlers r stündlich erhoben wurde und zur jeweiligen Produktgruppe g gehört. Der Buchstabe n gibt die Anzahl an Preisen an, die im Monat m auf der Internetseite gefunden wurden.

Die Verwendung des COICOP-Index hat den Vorteil, dass alle erhobenen Preisdaten verwendet werden können und mit entsprechenden Gewichten der Berichtsstellen nach Produktgruppe ein repräsentativer Markt abgebildet werden kann.

Allerdings werden hier alle erhobenen Preise implizit gleich gewichtet, unabhängig davon, wie oft beziehungsweise ob überhaupt zum jeweiligen Preis verkauft wurde.

Die Elementarindizes $I_{g,m}^{web\ scraping}$ der jeweiligen Produktgruppe, welche aus den Web-Scraping-Daten errechnet werden, können mit den korrespondierenden Elementarindizes $I_{g,m}^{CPI}$ verglichen werden. Diese werden monatlich gemeinsam von den Statistischen Ämtern der Länder und dem Statistischen Bundesamt im bisherigen Verfahren im Onlinehandel ermittelt und fließen in die laufende Produktion des VPI/HVPI ein. Gegenübergestellt werden also Indizes berechnet aus mit Web Scraping gewonnenen Massendaten und Indizes aus der regulären Berechnung des VPI/HVPI.

$$(6) \quad I_{g,m}^{web\ scraping} = \frac{P_{g,m}}{P_{g,m-1}}$$

$$(7) \quad I_{g,m}^{CPI} = \sum_{s=1}^{16} w_s \frac{P_{g,m}}{P_{g,m-1}}$$

Der Elementarindex $I_{g,m}^{CPI}$ entspricht dem gewichteten arithmetischen Mittel der Elementarindizes der einzelnen Bundesländer s für die jeweilige Produktgruppe g . Diese Elementarindizes werden mit Preisdaten berechnet, die von den Statistischen Ämtern der Länder und vom Statistischen Bundesamt im Onlinehandel erhoben werden und in die laufende Produktion des VPI/HVPI einfließen. Die Elementarindizes einer Produktgruppe liegen für die einzelnen Länder vor und werden entsprechend der Größe des Bundeslands gewichtet – w_s – und zu einem Ergebnis für Deutschland aggregiert. Die regulären (gesamtdeutschen) Elementarindizes werden mit den Elementarindizes verglichen, die sich aus der stündlichen (automatisierten) Preiserhebung ergeben.

Die vorläufigen Ergebnisse der experimentellen Elementarindizes seit Erhebungsbeginn im Dezember 2018 zeigen teilweise einen sehr unterschiedlichen Verlauf gegenüber den regulären Daten auf. Aus den bisher betrachteten Produktgruppen lässt sich jedoch schließen, dass bei häufigeren Preiserhebungen von Bekleidung im Onlinehandel die Entwicklung der Elementarindizes volatiler verlaufen kann als bei den regulären Daten. Betrachtet man beispielhaft die Elementarindizes aus den Studiendaten der Produktgruppen Herrenstoffhose und Herrenjeans, weisen diese zwar einen ähnlichen Verlauf auf, weisen jedoch deutliche Unterschiede auf.

10 Ein Elementarindex ergibt sich für jede Produktgruppe als Veränderungsrate der Durchschnittspreise der darin erhobenen Produkte zwischen aktuellem Monat und Vormonat.

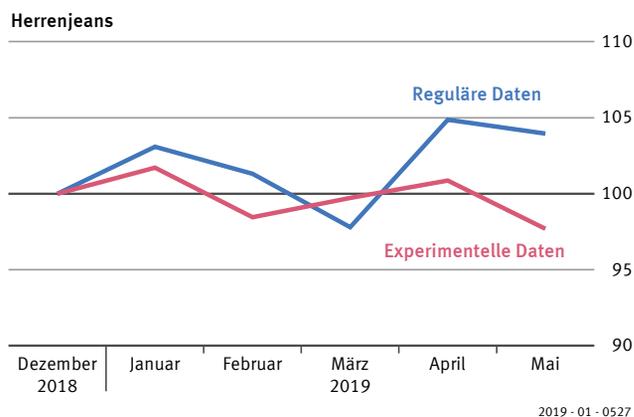
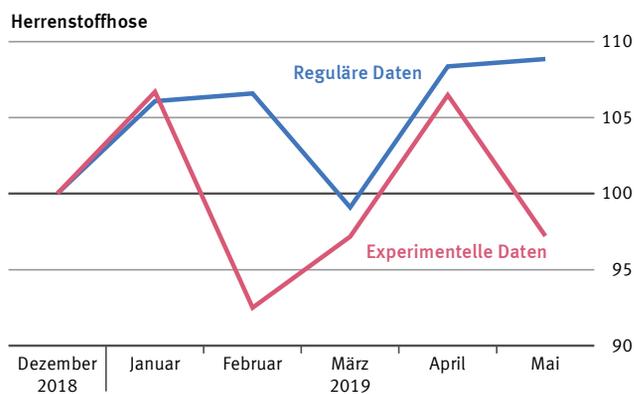
11 An diesem Ansatz wird allerdings ebenfalls intensiv geforscht. In früheren Studien wurde beispielsweise bereits festgestellt, dass die Anzahl der Preisänderungen auch abhängt vom Wochentag und der Tageszeit. Darüber hinaus ist anzunehmen, dass Konsumenten und Konsumentinnen ihre Einkäufe im Onlinehandel weniger häufig nachts tätigen.

lichen Verlauf, aber stärkere Preisschwankungen auf. Insbesondere beim Vergleich der Elementarindizes der Herrenstoffhose fällt auf, dass der Elementarindex nach regulären Daten dem Elementarindex nach experimentellen Daten zeitverzögert nachfolgt. Eine Erklärung für den zeitverzögerten Verlauf könnte sein, dass die experimentellen Daten bereits Preisänderungen enthalten, die die reguläre Erhebung aufgrund ihrer Ausgestaltung aktuell erst im Folgemonat berücksichtigen kann. Besonders auffällig ist dies in den Monaten Februar und März 2019: Bei der Herrenstoffhose handelt es sich um ein saisonales Produkt, bei dem zumeist in den Monaten Februar und März der Übergang von der Winter- zur Sommerkollektion vollzogen wird. Die Preiserhebungen sind in diesen beiden Monaten durch Sonderangebote geprägt. Der massive Abverkauf der Winterware und die Einführung der Sommerartikel zum regulären Preis werden durch die häufigeren Preiserhebungen zeitlich nach

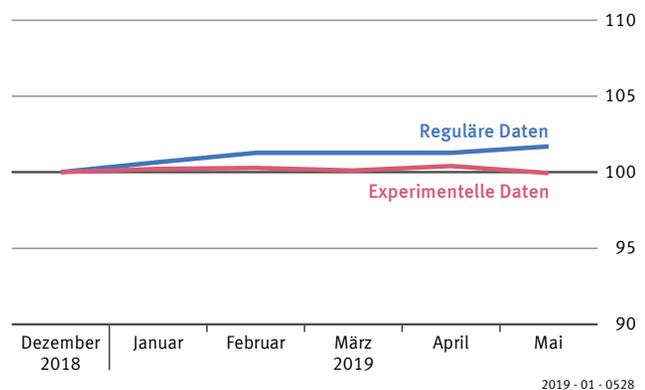
vorne verlagert. Im weiteren Verlauf der aktuellen Studien muss geprüft werden, ob es sich hierbei um eine noch zu korrigierende Verzerrung oder tatsächlich um eine bessere zeitliche Abdeckung handelt. Auch bei den Herrenjeans ist diese Verschiebung zu erkennen, wenn auch weit weniger ausgeprägt. [↘ Grafik 6](#)

Bei Produkten aus der Produktgruppe „Möbel, Leuchten, Geräte und anderes Haushaltszubehör“ (hier beispielhaft Pflegemittel) machen häufige Preiserhebungen offensichtlich kaum einen Unterschied für die Indexberechnung gegenüber der traditionellen Preiserhebung. [↘ Grafik 7](#)

Grafik 6
Vorläufige Ergebnisse der experimentellen Elementarindizes für die Beispiele Herrenstoffhose und Herrenjeans
Dezember 2018 = 100



Grafik 7
Vorläufige Ergebnisse der experimentellen Elementarindizes für das Beispiel Pflegemittel
Dezember 2018 = 100



Erweiterungen des COICOP-Index hinsichtlich umfassenderer Preisindizes im Onlinehandel sind vorstellbar. Die Analyse experimenteller Indexformeln und die Prüfung, ob diese für die deutsche Verbraucherpreisstatistik anwendbar sind, sind Bestandteil der laufenden Studie und werden in den folgenden Jahren weiter vorangetrieben.

8

Fazit und Ausblick

Den Herausforderungen durch die gestiegene Bedeutung des Internethandels begegnet die amtliche Preisstatistik auf zwei Arten:

1. Manuelle Preiserhebungen im Internet für die laufende Indexproduktion werden sukzessive automatisiert. Somit können diese Preiserhebungen künftig mit einer optimierten Anzahl an Wiederholungen durchgeführt werden. Auch die Stichprobe im Onlinehandel kann durch Web Scraping in kurzer Zeit ausgeweitet werden.
2. Derzeit werden neue Validierungsprozesse entwickelt und Berechnungsmethoden getestet, mit dem Ziel, auch sehr volatile Preise mittels großer Datenmengen besser abzubilden und in die monatliche Indexproduktion zu integrieren.

Die Einführung des generischen Programms in die Produktion des VPI/HVPI ist sukzessive ab der zweiten Jahreshälfte 2019 geplant. Mit dessen Hilfe wird die manuelle Preiserhebung im Internet schrittweise automatisiert beziehungsweise bestehende Automatisierungsprozesse in den generischen Standard überführt. Parallel werden die Preiserheberinnen und Preiserheber in der Anwendung des Programms geschult, um ihre Preiserhebungen eigenständig zu automatisieren und anschließend zu pflegen. Das Ziel ist, bis Ende des Jahres 2020 die vormals manuellen Preiserhebungen im Internet für 10 000 Produkte vollständig zu automatisieren. Zunächst werden dabei die bisherigen manuellen Erhebungen imitiert und bei volatilen Preisen die Erhebungsfrequenz angepasst. Methodische Änderungen sind in diesem ersten Schritt nicht vorgesehen. Erforderlich sind jedoch zusätzliche Validierungsprozesse während des Erhebungsprozesses.

Darüber hinaus werden neue Methoden untersucht und entwickelt, um die Möglichkeiten des Web Scrapings möglichst weitgehend auszuschöpfen. Das Ziel ist, zur nächsten Revision der Verbraucherpreisstatistik die oben beschriebenen neuen Berechnungsmethoden, soweit sinnvoll, einzubauen. Langfristig kann das generische Programm auch für die Datenerhebung anderer Statistiken erweitert und verwendet werden.

Die flächendeckende Integration von Web Scraping in die laufende Produktion der Preisstatistik stellt einen wichtigen Schritt hin zur Digitalisierung im Statistischen Bundesamt dar. Das generische Programm wird die Onlinepreiserhebungen deutlich verbessern und damit auch die Qualität des VPI und des HVPI steigern. 

LITERATURVERZEICHNIS

Bladow, Christian. *Fortschritte und Herausforderungen beim Web Scraping – Automatisierung von Preiserhebungen im Internet*. In: Methoden – Verfahren – Entwicklungen. Ausgabe 1/2018, Seite 3.

Bladow, Christian/Burg, Florian. *Dynamische Preissetzung als Herausforderung für die Verbraucherpreisstatistik*. In: WISTA Wirtschaft und Statistik. Ausgabe 2/2018, Seite 11 ff.

Bieg, Matthias. *Nutzung von Scannerdaten in der Preisstatistik – eine Untersuchung anhand von Marktforschungsdaten*. In: WISTA Wirtschaft und Statistik. Ausgabe 2/2019, Seite 25 ff.

Sandhop, Karsten. *Geschäftstypengewichtung im Verbraucherpreisindex*. In: Wirtschaft und Statistik. Ausgabe 3/2012, Seite 266 ff.

Burg, Florian/Seeger, Daniel. *Das neue Verbundprogramm der Verbraucherpreisstatistik*. In: Wirtschaft und Statistik. Ausgabe 2/2009, Seite 169 ff.

Griffioen, Robert/ten Bosch, Olav. *On the use of Internet data for the Dutch CPI*. 2016. [Zugriff am 12. Juni 2019]. Verfügbar unter: www.researchgate.net

Van Loon, Ken/Roels, Dorien. *Integrating big data in the Belgian CPI*. 2018. Meeting of the Group of Experts on Consumer Price Indices. Verfügbar unter: www.unece.org

Nygaard, Ragnhild. *The use of online prices in the Norwegian Consumer Price Index*. Tokio 2015.

Varian, Hal R. *Intermediate Microeconomics. A Modern Approach*. 8. Auflage. 2010.

DATENERGÄNZUNG DER STRUKTUR- ERHEBUNG ENERGIE MITHILFE EINES ROBUSTEN REGRESSIONSMODELLS

Maren Köhlmann

↳ **Schlüsselwörter:** Unternehmensstrukturstatistik – Verwaltungsdaten – robustes Regressionsmodell – Abschneidegrenze – statistisches Unternehmensregister

ZUSAMMENFASSUNG

Die wichtigste Quelle zur Darstellung der Unternehmenslandschaft und zur Beurteilung der strukturellen Entwicklung von Unternehmen in Deutschland sind die Strukturstatistiken. Sie liefern jährlich Daten zu Investitionen und zur Kostenstruktur verschiedener Wirtschaftszweige. Im Bereich der Energie-, Versorgungs- und Entsorgungswirtschaft basieren die Daten auf der Erhebung aller Unternehmen oberhalb bestimmter Abschneidegrenzen. Um jedoch die Struktur der Grundgesamtheit aller Wirtschaftseinheiten in diesem Bereich darstellen zu können, werden Verwaltungsdaten mithilfe des statistischen Unternehmensregisters hinzugespielt und hierauf aufbauend ein Datenergänzungsmodell erstellt. Dieser Beitrag stellt die methodische Konzeption dieser Datenergänzung, die Bedeutung der praktischen Implementierung und die Datenqualität dar.

↳ **Keywords:** *structural business statistics – administrative data – robust regression model – cut-off threshold – statistical business register*

ABSTRACT

Structural business statistics are the most important source for illustrating the corporate landscape and assessing the structural development of businesses in Germany. They provide annual data on investments and the cost structure in various economic sectors. In the energy, utilities and waste management sector, the data are based on a survey of all enterprises above certain cut-off thresholds. However, in order to depict the structure of the whole population of economic entities in this sector, administrative data from the business register are added and a data supplementation model is created on that basis. This article presents the methodological concepts for the supplementation of data, describes the importance of the practical implementation, and evaluates the data quality.



Maren Köhlmann

hat Internationale Volkswirtschaftslehre und Soziologie studiert. Sie ist im Referat „Energie, Ver- und Entsorgungswirtschaft“ des Statistischen Bundesamtes als wissenschaftliche Mitarbeiterin mit der methodischen Weiterentwicklung und Umsetzung von EU-Anforderungen der Struktur-erhebung in der Energie-, Wasser- und Entsorgungswirtschaft befasst.

1

Einleitung

Kostenstruktur- und Investitionsstatistiken der Unternehmen ermöglichen eine detaillierte Analyse der Unternehmenslandschaft. In der Bundesrepublik Deutschland werden Kostenstrukturerhebungen in der gewerblichen Wirtschaft seit dem Jahr 1959 jährlich durchgeführt.¹ Die Kostenstrukturerhebung ist ebenso wie die Investitionsenerhebung bei Unternehmen der Energieversorgung, Wasserversorgung, Abwasser- und Abfallentsorgung und zur Beseitigung von Umweltverschmutzungen (im Folgenden als „Strukturerhebung Energie“ bezeichnet) eine Primärerhebung mit Auskunftspflicht (Statistisches Bundesamt, 2018, hier: Seite 80). Diese Daten liefern wichtige Informationen über die Beschäftigtenstruktur, Tätigkeiten und Wettbewerbsfähigkeit der Unternehmen. Zudem fließen die Daten in die Volkswirtschaftlichen und Umweltökonomischen Gesamtrechnungen ein und sind auch auf europäischer Ebene von großer Bedeutung, beispielsweise zur Darstellung von Anteilen der Branche an der Gesamtwirtschaft. Strukturstatistiken sind daher von höchster Relevanz für Analysen und wirtschaftspolitische Entscheidungen.

Aufgrund der politischen und gesetzlichen Vorgabe, insbesondere kleinste und kleine Unternehmen bürokratisch zu entlasten², wird die Strukturerhebung Energie als Vollerhebung mit festgelegten Abschneidegrenzen durchgeführt. Diese Art der Erhebung, bei der alle Unternehmen oberhalb eines Grenzwerts (der sich beispielsweise auf die Beschäftigtenanzahl oder die Höhe des Umsatzes bezieht) auskunftspflichtig sind, wird auch als „Konzentrationsstichprobe“ bezeichnet (Dinges und andere, 2010, hier: Seite 282). Betreiber von Anlagen zur Erzeugung von Strom aus erneuerbaren Energien, die ihre Strommengen gemäß dem Erneuerbare-Energien-Gesetz³ in das Netz der allgemeinen Versorgung einspeisen, sind bisher nicht für die Strukturerhebung

1 Gesetz über Kostenstrukturstatistik (KoStrukStatG) vom 12. Mai 1959, in der im Bundesgesetzblatt Teil III, Gliederungsnummer 708-3, veröffentlichten bereinigten Fassung, das zuletzt durch Artikel 4 des Gesetzes vom 21. Juli 2016 (BGBl. I Seite 1768) geändert worden ist.

2 Siehe beispielsweise das Zweite Bürokratieentlastungsgesetz aus dem Jahr 2017.

3 Gesetz für den Ausbau erneuerbarer Energien (Erneuerbare-Energien-Gesetz – EEG 2017) vom 21. Juli 2014 (BGBl. I Seite 1066), das zuletzt durch Artikel 5 des Gesetzes vom 13. Mai 2019 (BGBl. I Seite 706) geändert worden ist.

Energie auskunftspflichtig, sofern ihr Jahresumsatz regelmäßig unter 5 Millionen Euro liegt (Statistisches Bundesamt, 2018, hier: Seite 80). Diese Abschneidegrenze wurde jedoch aufgrund geänderter Rahmenbedingungen ab dem Berichtsjahr 2018 angepasst (siehe Abschnitt 2.1).

Dem wirtschaftlichen Ziel, die Auskunftspflichtigen zu entlasten, stehen europäische Vorgaben gegenüber, die eine statistische Darstellung der Struktur der Grundgesamtheit verlangen. Da die Konzentrationsstichprobe kein repräsentatives Sample liefert und somit keine konventionelle Hochrechnung angewandt werden kann, besteht die Herausforderung darin, geeignete Methoden zur Darstellung der Grundgesamtheit zu realisieren. Um diesem Zielkonflikt zwischen Entlastung der Auskunftspflichtigen und Anforderungen zur Aussagekraft der Statistik zu begegnen, werden künftig die Daten der Strukturerhebung Energie mit Verwaltungsdaten aus dem statistischen Unternehmensregister verknüpft. Damit schließlich die Gesamtzahl der Unternehmen im Energie- sowie Versorgungs- und Entsorgungssektor dargestellt werden kann, wird basierend auf den durch Verwaltungsdaten ergänzten Unternehmensdaten ein Schätzmodell entwickelt. Dadurch ist es möglich, die weiteren Merkmale modellbasiert zu ergänzen.

Die folgenden Kapitel beschreiben zunächst die zugrunde liegende Datenstruktur der Erhebungs- und Verwaltungsdaten und erläutern anschließend das konzeptionelle Vorgehen für das Datenergänzungsmodell. Abschließend wird die Datenqualität des neuen ergänzten Datenbestands untersucht und ein Ausblick gegeben.

2

Datenbasis

Die Grundgesamtheit der Unternehmen im Energie-, Versorgungs- und Entsorgungssektor kann durch Daten aus dem statistischen Unternehmensregister dargestellt werden. Das statistische Unternehmensregister enthält jedoch keine Informationen über zusätzliche Strukturmerkmale wie die Bruttowertschöpfung oder Produktionswerte. Diese Angaben liefert die Strukturerhebung Energie für den Teil der Unternehmen, die

oberhalb der Abschneidegrenze liegen. Zur Generierung eines Gesamtdatenbestands, der die Strukturmerkmale fur die Gesamtpopulation der Unternehmen abbildet, wird daher zunachst eine Datenbasis erstellt, die die beiden Datenquellen der Erhebungsdaten und der Verwaltungsdaten verknupft. Auf Grundlage dieser Datenbasis wird ein Schatzverfahren entwickelt, das die Strukturmerkmale der unter der Abschneidegrenze der Erhebung liegenden statistischen Einheiten erganzt. Die beiden Datenquellen zur Erstellung der Datenbasis werden im Folgenden beschrieben.

2.1 Unternehmensstrukturdaten der Energieerhebung

Die wichtigste Quelle zur Darstellung der Unternehmenslandschaft in Deutschland sind die Strukturstatistiken. Sie liefern jahrliche Daten zu Investitionen und zur Kostenstruktur der Unternehmen verschiedener Wirtschaftszweige. Daher sind sie die zentrale Datenbasis zur Beurteilung der strukturellen Entwicklung von Unternehmen und eine elementare Grundlage fur detaillierte Analysen der Wirtschaftspolitik. Auerdem werden wichtige Daten aus den amtlichen Strukturhebungen fur die Volkswirtschaftlichen Gesamtrechnungen sowie fur umweltkonomische Statistiken genutzt. Im Bereich des Produzierenden Gewerbes werden die Strukturstatistiken auf der Rechtsgrundlage des Gesetzes ber die Statistik im Produzierenden Gewerbe (ProdGewStatG)⁴ in Verbindung mit dem Bundesstatistikgesetz erhoben. In den Abschnitten D „Energieversorgung“ und E „Wasserversorgung, Abwasser- und Abfallentsorgung, Beseitigung von Umweltverschmutzungen“ der Klassifikation der Wirtschaftszweige⁵ basieren die Daten wie zuvor beschrieben auf einer Vollerhebung oberhalb bestimmter Abschneidegrenzen. Die Werte fur diese Schwelle werden unter Berucksichtigung einer normativen Reprasentanzgrenze festgelegt. Dies bedeutet, dass die Anzahl der auskunftspflichtigen Unternehmen anhand von fachlichen Einschatzungen unter Beachtung der gesetzlichen Vorgaben ermittelt wird. Die nach § 6 und § 6a ProdGewStatG festgelegte Anzahl der einbe-

4 Gesetz ber die Statistik im Produzierenden Gewerbe (ProdGewStatG) in der Fassung der Bekanntmachung vom 21. Marz 2002 (BGBl. I Seite 1181), das zuletzt durch Artikel 271 der Verordnung vom 31. August 2015 (BGBl. I Seite 1474) geandert worden ist.

5 Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

zogenen Unternehmen betragt im Bereich der Energieversorgungsunternehmen aktuell maximal 3 000 Unternehmen. Bei den Unternehmen mit Schwerpunkt in der Wasserversorgung, Abwasser- und Abfallentsorgung, Beseitigung von Umweltverschmutzungen sind hchstens 7 000 Unternehmen zu befragen. [↘ bersicht 1 auf Seite 34](#)

Wie in Kapitel 1 beschrieben, wurden die Abschneidegrenzen fachlich neu bewertet und neu festgesetzt. Dies war notwendig geworden, da die bisherige enge Verzahnung der Berichtskreise der Strukturhebung Energie mit den Energiestatistiken aufgrund der strategischen Neuausrichtung der Energiestatistiken⁶ nicht mehr mglich war. Unter Berucksichtigung der gesetzlichen Vorgabe zur brokratischen Entlastung kleinster und kleiner Unternehmen werden beispielsweise im Bereich der Energieversorgung nunmehr Unternehmen befragt, deren Jahresumsatz 1 Million Euro (WZ 35.3) beziehungsweise 3 Millionen Euro (WZ 35.1 und 35.2) und mehr betragt. Ein anderes Kriterium als Abschneidegrenze wird zum Teil im Wirtschaftsabschnitt E verwendet, wo beispielsweise im Bereich der Wasserversorgung nur Unternehmen mit einer jahrlichen Wasserabgabe von 200 000 m³ und mehr herangezogen werden (Statistisches Bundesamt, 2018, hier: Seite 80).

2.2 Statistisches Unternehmensregister

Eine zweite Datenquelle zur Darstellung wichtiger Merkmale der Unternehmenslandschaft basiert auf Auswertungen aus dem statistischen Unternehmensregister. Auf rechtlicher Grundlage des Verwaltungsdatenverwendungsgesetzes⁷ speist sich das statistische Unternehmensregister aus Umsatzdaten der Finanzverwaltungen (Oberfinanzdirektionen) und Beschaftigtendaten der Bundesagentur fur Arbeit auf Einzeldatenebene. Fur jeden Datensatz werden Unternehmensumsatze aus den monatlichen Daten des Umsatzsteuervorauszahlungs- und Voranmeldungsverfahrens gewonnen. Das

6 Die strategische Neuausrichtung der Energiestatistiken wurde bedingt durch die Novelle des Energiestatistikgesetzes (Details zu den Energiestatistiken und der Novelle des Energiestatistikgesetzes beschreiben Decker/Klump, 2017).

7 Gesetz ber die Verwendung von Verwaltungsdaten fur Zwecke der Wirtschaftsstatistiken (Verwaltungsdatenverwendungsgesetz – VwDVG) vom 4. November 2010 (BGBl. I Seite 1480), das durch Artikel 1 des Gesetzes vom 18. Dezember 2018 (BGBl. I Seite 2637) geandert worden ist.

Übersicht 1

Klassifizierungen der Energiestatistiken nach der Statistischen Systematik der Wirtschaftszweige¹

Abschnitt D – Energieversorgung	
35 Energieversorgung	} maximal 3 000 Auskunfts- pflichtige
35.1 Elektrizitätsversorgung	
35.11 Elektrizitätserzeugung	
35.12 Elektrizitätsübertragung	
35.13 Elektrizitätsverteilung	
35.14 Elektrizitätshandel	
35.2 Gasversorgung	
35.21 Gaserzeugung	
35.22 Gasverteilung durch Rohrleitungen	
35.23 Gashandel durch Rohrleitungen	
35.3 Wärme- und Kälteversorgung	
35.30 Wärme- und Kälteversorgung	
Abschnitt E – Wasserversorgung; Abwasser- und Abfallentsorgung und Beseitigung von Umweltverschmutzungen	
36 Wasserversorgung	} maximal 7 000 Auskunfts- pflichtige
36.0 Wasserversorgung	
36.00 Wasserversorgung	
37 Abwasserentsorgung	
37.0 Abwasserentsorgung	
37.00 Abwasserentsorgung	
38 Sammlung, Behandlung und Beseitigung von Abfällen; Rückgewinnung	
38.1 Sammlung von Abfällen	
38.11 Sammlung nicht gefährlicher Abfälle	
38.12 Sammlung gefährlicher Abfälle	
38.2 Abfallbehandlung und -beseitigung	
38.21 Behandlung und Beseitigung nicht gefährlicher Abfälle	
38.22 Behandlung und Beseitigung gefährlicher Abfälle	
38.3 Rückgewinnung	
38.31 Zerlegen von Schiffs- und Fahrzeugwracks und anderen Altwaren	
38.32 Rückgewinnung sortierter Werkstoffe	
39 Beseitigung von Umweltverschmutzungen und sonstige Entsorgung	
39.0 Beseitigung von Umweltverschmutzungen und sonstige Entsorgung	
39.00 Beseitigung von Umweltverschmutzungen und sonstige Entsorgung	

¹ Eurostat. NACE Rev. 2. Statistische Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft, Seite 74 f. NACE ist das Akronym aus „Nomenclature statistique des activités économiques dans la Communauté européenne“. 2008. National: Statistisches Bundesamt. Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008). 2008.

statistische Unternehmensregister enthält alle Unternehmen (rechtliche Einheiten), die einen kumulierten steuerbaren Jahresumsatz oberhalb der Grenze von 17 500 Euro haben (Lorenz/Opfermann, 2017, hier: Seite 53). Die Daten umfassen neben den Umsatz- und Steuerpositionen des Umsatzsteuervorauszahlungs- und Voranmeldungsverfahrens ebenso Ordnungsmerkmale, beispielsweise Wirtschaftszweig, Rechtsform und Gemeindeschlüssel. Die Beschäftigtendaten für das statistische Unternehmensregister werden aus den monatlichen betriebsbezogenen Angaben zu sozialversiche-

rungspflichtig Beschäftigten und geringfügig entlohnt Beschäftigten der Bundesagentur für Arbeit gewonnen. Auch diese Daten umfassen die oben genannten Ordnungsmerkmale bezogen auf die Betriebe. Für den verknüpften Datenbestand werden die Daten der sozialversicherungspflichtig Beschäftigten und die der geringfügig entlohnt Beschäftigten addiert, wie in Kapitel 3 erläutert wird.

2.3 Herausforderungen bei der Nutzung von Daten aus dem statistischen Unternehmensregister

Bei der Nutzung von Daten aus verschiedenen Quellen ist darauf hinzuweisen, dass sich die Daten im Hinblick auf bestimmte Definitionsabgrenzungen unterscheiden (Lorenz/Opfermann, 2017). Dies trifft auch auf die Primardatennutzung aus der Strukturhebung Energie und der sekundaren Datenquelle der Verwaltungsdaten aus dem statistischen Unternehmensregister zu. In der Erhebung umfasst der Gesamtumsatz alle Lieferungen und Leistungen (ohne Umsatzsteuer, ohne Stromsteuer, ohne Erdgassteuer), die Dritten in Rechnung gestellt werden. Auerordentliche und betriebsfremde Ertrage werden hierbei nicht miteinbezogen, wohingegen diese in den an die Finanzverwaltung gemeldeten Umsatzdaten enthalten sind. Weitere definitorische Unterschiede, die sich ebenso auf die Beschaftigtendaten beziehen, wurden in dieser Zeitschrift bereits ausfuhrlich erortert (Lorenz/Opfermann, 2017, sowie Dechent, 2017, hier: Seite 64 f.). Eine Gegenuberstellung der wichtigsten definitorischen Unterschiede enthalt [bersicht 2](#).

Bei den unterschiedlichen Verwaltungsdatenquellen kann es zudem zu verschiedenen Zuordnungen der Unternehmen oder Betriebe zu den jeweiligen Wirtschaftszweigen kommen, weil die Zuordnung nicht einheitlich ber eine zentrale Institution erfolgt. Somit knnten verschiedene angegebene wirtschaftliche Tatigkeitsschwerpunkte der Unternehmen unterschiedlich klassifiziert werden. Dieser Aspekt ist fur die Datenbasis des Schatzmodells von geringer Bedeutung, da die Daten aus der Finanzverwaltung und der Bundesagentur

fur Arbeit im statistischen Unternehmensregister bereits zusammengefuhrt werden. Die verknpfte Datenbasis fur das Schatzmodell im Fall der Strukturdaten im Energiebereich basiert auf der im statistischen Unternehmensregister zugrunde gelegten Wirtschaftszweiguordnung.

Neben definitorischen Unterschieden oder Zuordnungsabweichungen spielen auch die Zeitpunkte der Datenerhebung und Registerpflege eine groe Rolle: Bei der primarstatistischen Erhebung wird der Berichtskreis zu einem bestimmten Zeitpunkt (Stichtag) im Jahr erstellt. Dagegen findet die Aktualisierung der Informationen des Unternehmensregisters anhand unterjahriger und jahrlicher Informationen der ffentlichen Verwaltung sowie aus den einzelnen primarstatistischen Erhebungen statt. Durch diese laufende Pflege knnen sich bei der Analyse der Daten Abweichungen ergeben, sodass eine exakte bereinstimmung der beiden Quellen nicht immer gegeben ist.

3

Verknpfung der Datenquellen und Eckdatenerganzung

Generell ergibt die Verknpfung von Einzeldaten aus verschiedenen Statistiken einen Mehrwert, da hierdurch neue Informationen gewonnen werden knnen, ohne Auskunftspflichtige weiter zu belasten (Jung/Kuser, 2016, hier: Seite 95). Die Verknpfung der Einzelbeobachtungen der beiden zuvor beschriebenen Datenquellen entspricht genau diesem Informationsgewinn. Durch die Zusammenfuhrung der Daten aus dem statistischen

bersicht 2

Gegenuberstellung der wichtigsten definitorischen Unterschiede von Primar- und Verwaltungsdaten

Erhebungsdaten	Verwaltungsdaten aus dem statistischen Unternehmensregister
Umsatz: <ul style="list-style-type: none"> > enthalt steuerfreie Umsatze, Handelsumsatze und Binnenhandel (Lieferungen und Leistungen innerhalb eines Konzerns) > enthalt keine auerordentlichen und betriebsfremden Ertrage > umfasst den Gesamtbetrag der Dritten in Rechnung gestellten Lieferungen und Leistungen (ohne Umsatzsteuer) 	Umsatz: <ul style="list-style-type: none"> > steuerfreie Umsatze und Binnenhandel werden bei umsatzsteuerlichen Organschaften nicht einbezogen > enthalt auerordentliche und betriebsfremde Ertrage
Beschaftigte („tatige Personen“): <ul style="list-style-type: none"> > Summe der tatigen Inhaber, der unbezahlt mithelfenden Familienangehrigen und der abhangig Beschaftigten > enthalten sind Erkrankte, im Urlaub, im Mutterschutz oder Erziehungsurlaub befindliche Personen und sonstige vorubergehend Abwesende sowie Saison- und Aushilfspersonen, Teilzeitbeschaftigte und Kurzarbeitende 	Beschaftigte: <ul style="list-style-type: none"> > umfasst die sozialversicherungspflichtig Beschaftigten und die geringfugig entlohnnten Beschaftigten > nicht enthalten in den Daten der Bundesagentur fur Arbeit sind tatige Inhaber, nicht sozialversicherungspflichtige Gesellschafter, mithelfende Familienangehrige sowie kurzfristig geringfugig Beschaftigte

Unternehmensregister mit denen der Strukturhebung Energie erhält man eine erweiterte Datenbasis. Diese stellt einerseits die Grundgesamtheit dar und enthält andererseits für die jeweils befragten Einheiten zusätzlich die Informationen der Strukturmerkmale. Die Verknüpfung der beiden Datenbestände erfolgt anhand der in beiden Quellen vorhandenen Unternehmensnummer des statistischen Unternehmensregisters. Durch diese in beiden Datenbeständen identische Unternehmensnummer kann jeder Einzeldatensatz exakt identifiziert und mit den weiteren Merkmalen zusammengeführt werden.

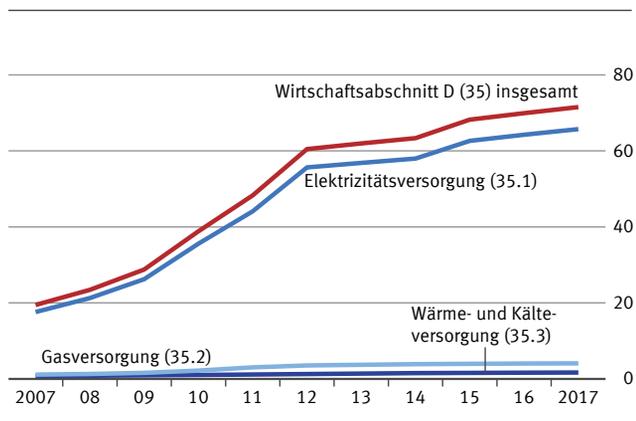
Die für das Berichtsjahr 2017 auf diesem Weg erstellte Gesamtdatenbasis besteht aus etwa 80 000 Einheiten¹⁸ für die Wirtschaftszweige D und E. Diese Zahl schwankt je nach Erhebungsjahr und hat in den vergangenen Jahren leicht zugenommen. Grund für die steigende Anzahl vor allem kleinster und kleiner Unternehmen ist neben den Auswirkungen des Unbundling (siehe Exkurs) und gesetzlicher Liberalisierungen im Energiesektor auch die steigende Anzahl der Unternehmen im Sektor der Erneuerbaren Energien.

Unbundling im Energiesektor

Bereits im Dezember 1996 wurde mit der Richtlinie 96/92/EG des Europäischen Parlaments und des Rates betreffend gemeinsame Vorschriften für den Elektrizitätsbinnenmarkt (Amtsblatt der EG Nr. L 027 vom 30. Januar 1997, Seite 20) ein erster Schritt für einen europäischen Elektrizitätsbinnenmarkt sowie für eine Entflechtung (sogenanntes Unbundling) der Energieversorgungsunternehmen getan (Bayer, 2011, hier: Seite 754). Unbundling beschreibt die gesetzliche Forderung nach einer Trennung des Netzbetriebs von anderen Tätigkeiten – wie beispielsweise dem Vertrieb – der Energieversorgungsunternehmen, um somit Transparenz und Wettbewerb am Markt zu fördern (Bundesnetzagentur, 2019). Auch wenn in Deutschland gewisse Unternehmen von dieser Regelung ausgenommen sind, führte das Unbundling von Netzbetrieb und Energieversorgung ebenso wie eine gesetzlich verordnete Liberalisierung zu einer veränderten Struktur des Energiemarktes (Decker/Klumpp, 2017, hier: Seite 64). Infolgedessen stieg die Anzahl der Unternehmen im Bereich der Energieversorgung an. [↘ Grafik 1](#)

¹⁸ Eine rechtliche Einheit beschreibt ein Unternehmen. Ein Unternehmen kann dabei aus mehreren Betrieben bestehen. Nach dem Gesetz über die Statistik im Produzierenden Gewerbe werden Betriebe und Unternehmen der Energieversorgung (§ 6) und der Wasserversorgung, der Abwasser- und Abfallentsorgung und der Beseitigung von Umweltverschmutzungen (§ 6a) befragt.

Grafik 1
Entwicklung der Unternehmenszahl
im Wirtschaftsabschnitt D – Energieversorgung
1 000



Nach der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

2019 - 01 - 0512

In beiden Datenquellen geführte Merkmale, bei denen für die Einheiten unterhalb der Abschneidegrenze die Angaben aus den Verwaltungsdaten übernommen werden, werden als Eckdaten bezeichnet. Auf diesen Eckdaten beruhen die späteren Schätzungen. Der neu erstellte verknüpfte Datenbestand enthält drei dieser Eckdatenmerkmale: die zur Verknüpfung genutzte Unternehmensnummer, Unternehmensumsätze und die Anzahl der Beschäftigten. Liegen für Einheiten Erhebungsdaten vor, sind diese immer zu nutzen. Wie in Abschnitt 2.3 erläutert, stimmen die Definitionen und Merkmalsbezeichnungen in den verschiedenen Datenquellen nicht überein. Das Eckdatenmerkmal „Beschäftigte“ im verknüpften Datenbestand wird aus dem Merkmal „Gesamtzahl der tätigen Personen“ der Erhebungsdaten und den addierten Werten von sozialversicherungspflichtig Beschäftigten und geringfügig entlohnten Beschäftigten der Registerdaten erstellt. Dennoch bleiben bei diesem Merkmal einzelne Lücken, nicht jeder Einzeldatensatz ist mit einem Wert abgedeckt. Dies ist möglich, weil manche Kleinstunternehmen keine Beschäftigten aufweisen. Dieser Aspekt ist bei der Zuschätzung zu berücksichtigen.

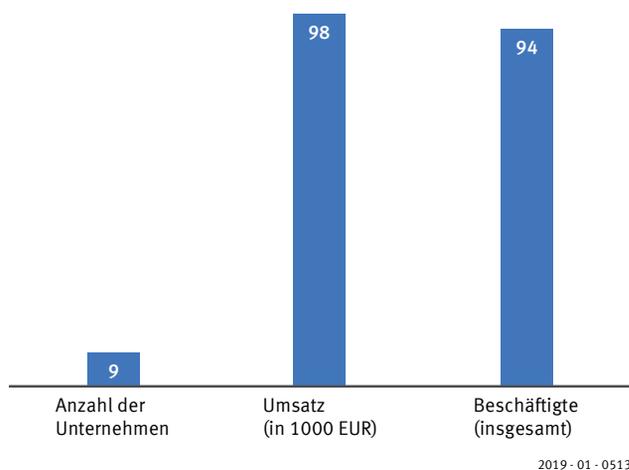
Die Analyse der erstellten Eckdaten der beiden Datenquellen zur Abbildung der Unternehmenslandschaft in Deutschland lässt Folgendes erkennen: Aufgrund der Abschneidegrenze besteht eine große Lücke zwischen der Anzahl der erhobenen Unternehmen und den in den Verwaltungsdaten erfassten Unternehmen. Trotzdem

sind die Merkmale Umsatz und Beschäftigtenanzahl durch Primärdaten fast vollständig abgedeckt. So mussten im Berichtsjahr 2017 nur rund 9% der über 80000 Unternehmen der Grundgesamtheit erhoben werden, um primärstatistisch 98% der Umsätze und 94% der Beschäftigten abbilden zu können. [↘ Grafik 2](#)

Grafik 2

Abdeckung der Merkmale Umsatz und Beschäftigte durch die Primärdaten der Strukturerhebung Energie 2017

in %



Das Konzept der Vollerhebung mit Abschneidegrenze führt somit dazu, dass vor allem Kleinstunternehmen⁹ von der Erhebung nicht erfasst werden, deren Umsatz

9 Aufteilung der Unternehmensgröße nach Umsatz in Anlehnung an die Umsatzgrößenklassen der Empfehlung (2003/361/EG) der Europäischen Kommission vom 6. Mai 2003 (Amtsblatt der EU Nr. L124, Seite 36 vom 20. Mai 2003):

- 1) Kleinstunternehmen: bis 2 Millionen Euro Umsatz
- 2) Kleine Unternehmen: bis 10 Millionen Euro Umsatz
- 3) Mittlere Unternehmen: bis 50 Millionen Euro Umsatz
- 4) Große Unternehmen: mehr als 50 Millionen Euro Umsatz

Table 1

Unternehmen im Wirtschaftszweig Elektrizitätsversorgung (WZ 35.1) 2017

	Unternehmen insgesamt	Kleinstunternehmen	Kleine Unternehmen	Mittlere Unternehmen	Große Unternehmen
35111 Elektrizitätserzeugung ohne Verteilung	5 938	5 902	15	10	11
35112 Elektrizitätserzeugung mit Fremdbezug zur Verteilung	11 784	11 536	47	107	94
35113 Elektrizitätserzeugung ohne Fremdbezug zur Verteilung	43 582	43 435	52	56	39
35120 Elektrizitätsübertragung	351	343	0	0	8
35130 Elektrizitätsverteilung	1 182	705	111	195	171
35140 Elektrizitätshandel	1 015	639	79	146	151

Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008).

und Beschäftigtenanzahl nur einen sehr geringen Einfluss auf die Gesamtergebnisse der Unternehmenslandschaft im Bereich der Energieversorgung haben. Eine Analyse nach Umsatzgrößenklassen lässt zudem erkennen, dass die meisten dieser Unternehmen ihre wirtschaftliche Tätigkeit in der Elektrizitätserzeugung (WZ 35113 „Elektrizitätserzeugung ohne Fremdbezug zur Verteilung“) haben. Gerade in dieser WZ-Klasse finden sich seit einigen Jahren sehr viele Kleinstunternehmen mit Stromerzeugungsanlagen, die aus erneuerbaren Energien gespeist werden, beispielsweise Photovoltaik- oder Solarthermie-Anlagen und Windräder. [↘ Tabelle 1](#)

Zur Eckdatenerstellung werden bei den Unternehmen, für die Primärdaten vorliegen, für den verknüpften Datenbestand die Primärdaten übernommen. Für diejenigen Unternehmen, für die keine Primärdaten vorliegen, werden die beiden Merkmale Umsatz und Beschäftigte aus den vorhandenen Verwaltungsdaten ergänzt. Für die weiteren Merkmale, die in der Primärerhebung abgefragt werden, gibt es allerdings keine äquivalenten Daten aus Sekundärquellen. Folglich wird anhand des verknüpften Datenbestands, der die Grundgesamtheit der Unternehmen enthält, mit statistischen Schätzverfahren eine Ergänzung der fehlenden Merkmale vorgenommen.

4

Merkmalsergänzung anhand von Schätzmodellen

Nach der Eckdatenerstellung werden die Strukturmerkmale der unterhalb der Abschneidegrenze liegenden Einheiten anhand von Regressionsmodellen geschätzt.

Diese Schätzungen basieren auf den Verwaltungsdaten der Einheiten unterhalb der Abschneidegrenze und den erhobenen Daten der Einheiten oberhalb der Abschneidegrenze. Es wird ein lineares Regressionsmodell zugrunde gelegt, weil die meisten wirtschaftsstatistischen Daten durch ein einfach lineares Modell gut beschrieben werden können. Dies gilt auch für die Daten der Strukturerhebung Energie, wie weiter unten am Beispiel des Merkmals Arbeitsstunden grafisch dargestellt ist (siehe Grafik 4). Das einfachste Modell in diesem Fall stellt die Regression der kleinsten Quadrate (Ordinary Least Squares = OLS) dar. Jedoch sollte bei der Merkmalsergänzung beachtet werden, dass wirtschaftsstatistische Daten häufig Ausreißer enthalten, die der Modellannahme der Linearität widersprechen. Um dieser Tatsache gerecht zu werden und eine stabile Qualität der Modellanpassung der Daten zu erreichen, wird eine robuste Technik angewandt.

Die Merkmalsergänzung auf der Ebene der Einzeldaten (Mikrodatenebene) erfolgt daher durch die Methode einer robusten linearen Regressionsschätzung. Um möglichst genaue Ergebnisse zu erzielen ist darauf zu achten, dass eine große Übereinstimmung der Struktur der Unternehmen vorliegt. Da sich die Unternehmen in dem

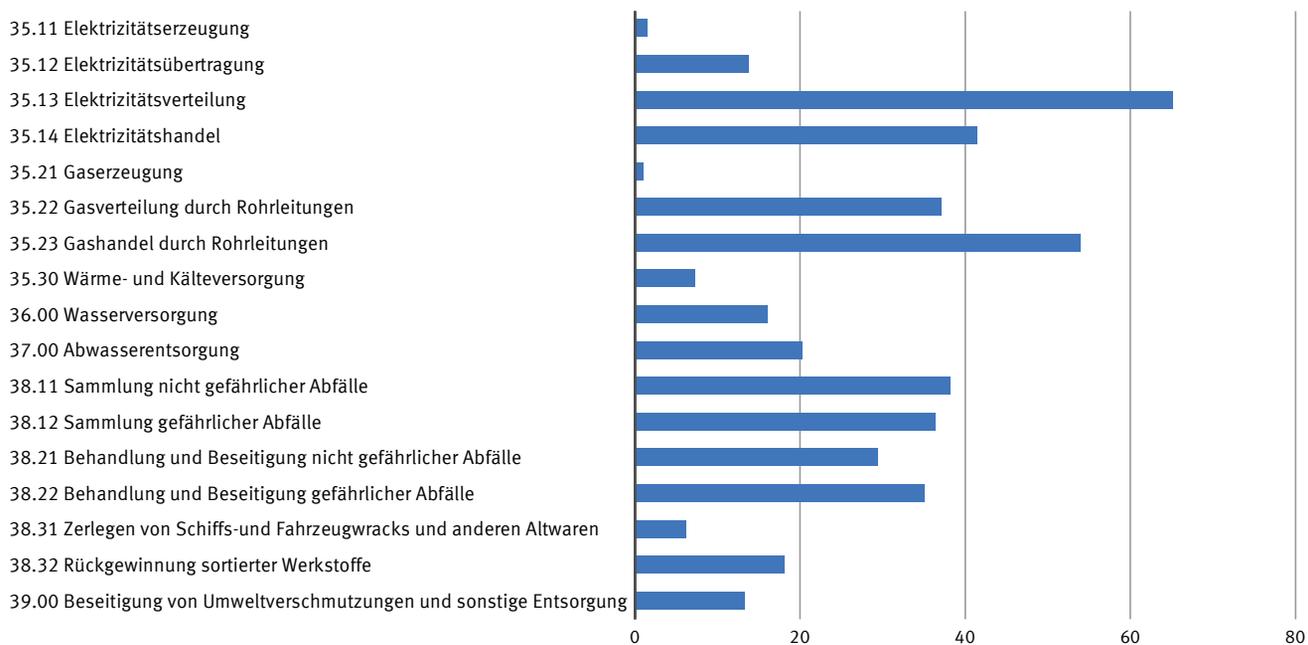
verknüpften Gesamtdatenbestand allerdings in ihrer Struktur unterscheiden können, werden für die Regressionsschätzungen Untergruppen nach Wirtschaftsbe-
reichen gebildet (siehe Übersicht 1). [Grafik 3](#) veranschaulicht, dass beispielsweise die Größe der Unternehmen gemessen an der Beschäftigtenanzahl in den verschiedenen Wirtschaftszweigen sehr heterogen ist: Während in manchen Untergruppen die Beschäftigtenanzahl bei unter einer Person je Unternehmen liegt, beträgt diese zum Beispiel in der Klasse der Elektrizitätsverteilung (WZ 35.13) rund 65 Personen.

Für die Regressionen werden also möglichst ähnliche Einheiten zugrunde gelegt, indem die Daten in die tiefste Gliederung der Wirtschaftszweige unterteilt werden, die für eine Regression die notwendige Datenbasis bereitstellt. So werden branchenspezifische Unterschiede der Unternehmen bei der Erstellung der Ergebnisse berücksichtigt. Probleme können auftauchen, wenn die Unterteilung der Unternehmen in Gruppen zu einer zu kleinen Datenbasis führt, sodass die Methode keine zuverlässigen Schätzer mehr liefert.

In der Literatur wird als eine der wahrscheinlich am meisten genutzten robusten Methoden die Least-Trim-

Grafik 3

Exemplarisches Beispiel für Strukturunterschiede der Unternehmen verschiedener Wirtschaftszweige
Beschäftigte je Unternehmen



2019 - 01 - 0515

Least-Squares (LTS)-Schatzung nach Rousseeuw (1984) diskutiert (Alfons und andere, 2013, hier: Seite 227; Dinges und andere, 2010, hier: Seite 290). Die LTS-Schatzung zeigt bestandige Ergebnisse bei der Prasenz von Ausreißern in den Daten, da sie den Einfluss dieser begrenzt (Gürünlü Alma, 2011, hier: Seite 410). Für die Regressionsgleichung

$$(1) \quad y_i = \beta_i x_i + \varepsilon_i,$$

bei der y_i die abhängige Variable, β_i den Regressionskoeffizienten, x_i die unabhängige beziehungsweise beobachtete Variable und ε_i den Fehlerterm darstellt, ist der LTS-Schatzter nach Rousseeuw (1984) wie folgt konstruiert:

$$(2) \quad \hat{\beta}_{LTS} = \arg \min_{\beta} Q_{LTS}(\beta), \text{ wobei } Q_{LTS}(\beta) = \sum_{i=1}^h e_i^2.$$

Hierbei sind $e_1^2 \leq e_2^2 \leq \dots \leq e_n^2$ die geordneten Residuen, $h = \left(\frac{n}{2}\right) + \left[\frac{p+1}{2}\right]$ ist der Umfang der Beobachtungen innerhalb der jeweils zugrunde liegenden Wirtschaftszweigaufteilung. Die LTS-Schatzung ist eine iterative Methode, bei der zufallig eine Teilmenge von Einheiten aus der Gesamtzahl der Einheiten gezogen wird. Aus dieser Teilmenge wird die Least-Squares-Regression und die Summe der kleinsten quadratischen Residuen berech-

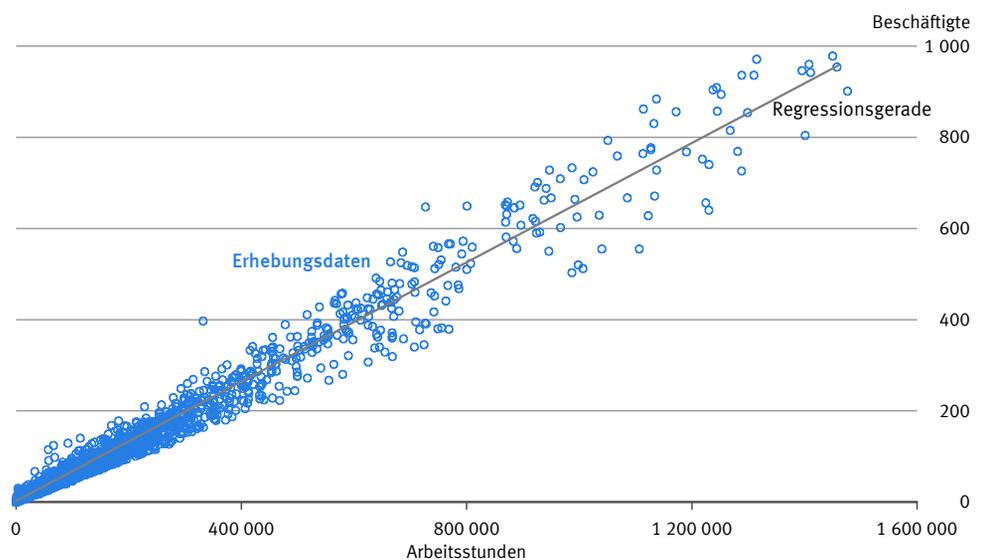
net (Dinges und andere, 2010, hier: Seite 291). Der LTS-Schatzter entspricht der Anpassung jener Teilmenge, bei der die kleinsten quadratischen Residuen minimiert sind. Grundlage für die Berechnung der zu erganzenden Merkmale der Erhebungsdaten für die Grundgesamtheit sind die jeweiligen wirtschaftszweigspezifischen Regressionsparameter. Die erganzten Merkmale basieren somit auf den zuvor erstellten Eckdaten Umsatz und Beschäftigtenanzahl.

Als Beispiel für die Ergebnisdarstellung wird im Folgenden die Anpassung für das Merkmal Arbeitsstunden herangezogen. Die Erganzung dieses Merkmals findet beispielsweise auf Grundlage der Beschäftigtenzahl nach oben beschriebener Methode statt:

$$(3) \quad \text{Arbeitsstunden} = \beta_i \text{Beschäftigte}_i + \varepsilon_i$$

In [Grafik 4](#) sind die Datenpunkte der Erhebungsdaten und die angepasste Regressionsgerade dargestellt. Weil das Eckdatenmerkmal Beschäftigte nicht für alle Einheiten des Datenbestands vorliegt, findet eine größere Unterteilung der Unternehmen in fünf Untergruppen für die Regressions-schatzungen auf WZ-2-Steller-Ebene für das Erhebungsmerkmal Arbeitsstunden statt. Daher enthalt das Ergebnis der in Gleichung (3) beschriebenen Regression funf Parameterschatzer (β_i). Das Gütemaß R^2 für diese gruppierten Regressionen liegt zwischen

Grafik 4
Lineare Regressionsanpassung von Arbeitsstunden zu Beschäftigten



2019 - 01 - 0516

Tabelle 2

Parameterschätzer und Regressionsgleichungen des Erhebungsmerkmals Arbeitsstunden für das Berichtsjahr 2017

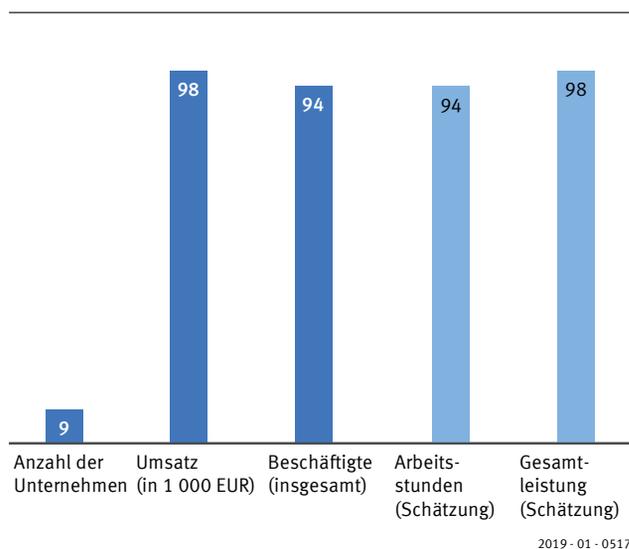
WZ-2-Steller	Regressionschätzung Arbeitsstunden	Gütemaß R ²
35 Energieversorgung	$Arbeitsstunden = 1\,497,93; Beschäftigte_j + \varepsilon_j$	0.9913
36 Wasserversorgung	$Arbeitsstunden = 1\,459,34; Beschäftigte_j + \varepsilon_j$	0.9687
37 Abwasserentsorgung	$Arbeitsstunden = 1\,489,45; Beschäftigte_j + \varepsilon_j$	0.9831
38 Sammlung, Behandlung und Beseitigung von Abfällen; Rückgewinnung	$Arbeitsstunden = 1\,523,54; Beschäftigte_j + \varepsilon_j$	0.9834
39 Beseitigung von Umweltverschmutzungen und sonstige Entsorgung	$Arbeitsstunden = 1\,568,37; Beschäftigte_j + \varepsilon_j$	0.9902

97 und 99% für das Merkmal Arbeitsstunden, was eine sehr gute Modellanpassung beschreibt. [↪ Tabelle 2](#) Jedoch ist zu bemerken, dass nicht für alle ergänzten Merkmalsdaten so hohe Gütemaße mit den Anpassungen erzielt werden können.

Es ist zu erkennen, dass beispielsweise in der Energieversorgung (WZ 35) je Unternehmen von einer zusätzlich beschäftigten Person etwa 1 498 Stunden im Jahr mehr gearbeitet werden (Berechnung für Daten des Berichtsjahres 2017). Außerdem zeigen die Ergebnisse, dass im Wirtschaftszweig E verhältnismäßig mehr Stunden je beschäftigter Person getätigt werden. Im Gegensatz zu der oben dargestellten Merkmalschätzung kann für die meisten Erhebungsmerkmale, bei denen die Schätzung basierend auf den Umsatzeckdaten erfolgt, die Gruppierung auf WZ-3-Steller-Ebene vorgenommen werden.

Grafik 5

Abdeckung der Merkmale Arbeitsstunden und Gesamtleistung durch Schätzungen für die Strukturhebung Energie 2017 in %



Wie zuvor beschrieben kann der Datenbestand der Grundgesamtheit durch die Regressionsschätzungen um die Erhebungsmerkmale ergänzt werden. Für den Großteil der Einheiten der Grundgesamtheit ist somit eine Abdeckung der Merkmale durch Schätzungen zu erreichen. Die Abdeckungsergebnisse für die Modellanpassungen der Merkmale Arbeitsstunden und Gesamtleistung sind exemplarisch dargestellt in [↪ Grafik 5](#).

5

Datenqualität der Merkmalsergänzung

Durch die Merkmalsergänzung werden die primärstatistisch erfassten Daten für den Teil der Unternehmen, die unterhalb der Abschneidegrenze liegen, ergänzt. Dadurch kann die Struktur der Grundgesamtheit aller Unternehmen in den Wirtschaftsabschnitten D „Energieversorgung“ und E „Wasserversorgung; Abwasser- und Abfallentsorgung, Beseitigung von Umweltverschmutzungen“ vollständig abgebildet werden.

Die in Kapitel 4 beschriebene robuste Regressionsmethode zur Ergänzung der fehlenden Merkmale aus der Erhebung für den Gesamtdatenbestand liefert insgesamt gute Ergebnisse und realistische Daten auf Mikrodatenebene. Es ist jedoch zu beachten, dass die Schätzungen immer dann zu Problemen führen, wenn in bestimmten Wirtschaftsbereichen keine ausreichende primärstatistische Datenbasis vorhanden ist. In diesem Fall wird die Untergliederung der Wirtschaftszweige weniger tief vorgenommen, um somit die Datenbasis entsprechend zu erweitern. Liegen starke branchenspezifische Strukturunterschiede vor, können diese in größer untergliederten Wirtschaftszweigen nicht mehr gut abgebildet werden. Dieser Zielkonflikt ist bei den ergänzten Daten immer zu beachten, sofern diese für weitere Analysen und Rechnungen herangezogen werden.

6

Ausblick

Die Verknüpfung von primärstatistisch erhobenen Daten der Strukturhebung Energie mit den sekundärstatistischen Verwaltungsdaten aus dem statistischen Unternehmensregister lässt die Datenergänzung der Strukturmerkmale anhand einer robusten Regressionsschätzung zu. Durch diese Datenergänzung wird der Gesamtdatenbestand der Unternehmen der Wirtschaftsabschnitte D und E vollständig auf Mikrodatenebene abgebildet. Auf dieser Grundlage können die bestehenden europäischen Meldepflichten komplett erfüllt werden. Dies war bisher aufgrund der fehlenden, unter der Abschneidegrenze liegenden Einheiten nicht gesichert. Darüber hinaus liefert der ergänzte Gesamtdatenbestand auch für die Statistiken, welche die Daten der Strukturhebung Energie nutzen, weitere Qualitätsgewinne. Ebenso bietet der verknüpfte Einzeldatenbestand große Potenziale für weitere Forschungszwecke.

Allgemein bietet die Nutzung verschiedener Datenquellen durch eine Datenverknüpfung auf Mikrodatenebene enorme Potenziale für die amtliche Statistik. Zum einen kann es bei Vorliegen von umfassenden Informationen aus Verwaltungsquellen zu Entlastungen von Auskunftspflichtigen kommen, zum anderen kann die Datenqualität durch die Auswertung verschiedener Quellen gesteigert werden. Zwar liegen aktuell für die Strukturhebung Energie keine weiteren Datenquellen zur Nutzung vor. Dieses Schätzmodell eröffnet aber die Möglichkeit, die künftige Nutzung weiterer im Aufbau befindlicher Register – beispielsweise das Marktstammdatenregister der Bundesnetzagentur oder ein mögliches Basisregister über Unternehmen – intensiv zu prüfen. 

LITERATURVERZEICHNIS

Alfons, Andreas/Croux, Christophe/Gelper, Sarah. *Sparse Least Trimmed Squares Regression for Analyzing High-Dimensional Large Data Sets*. In: The Annals of Applied Statistics. 2013. Jahrgang 7. Nr. 1, Seite 226.

Bayer, Wolfgang. [Sich ständig wandelnde Energiemärkte – eine Herausforderung für die amtliche Energiestatistik](#). In: Wirtschaft und Statistik. Ausgabe 8/2011, Seite 753 ff.

Bundesnetzagentur (BNetzA). *Entflechtung, Konzessionen, Geschlossene Verteilernetze*. [Zugriff am 2. Juli 2019]. Verfügbar unter: www.bundesnetzagentur.de

Dechent, Jens. [Die Mixmodelle in den Konjunkturstatistiken des Bauhaupt- und Ausbaugewerbes](#). In: WISTA Wirtschaft und Statistik. Ausgabe 3/2017, Seite 61 ff.

Decker, Jörg/Klumpp, Nicolai. [Strategische Neuausrichtung der Energiestatistiken](#). In: WISTA Wirtschaft und Statistik. Ausgabe 2/2017, Seite 63 ff.

Dinges, Gerlinde/Haitzmann, Martin/Zach, Sabine. *Methodenbericht zur Modellbasierten Datenergänzung in der Leistungs- und Strukturstatistik*. In: Austrian Journal of Statistics. 2010. Jahrgang 39, Nr. 4, Seite 281 ff.

Eurostat. *NACE Rev. 2. Statistische Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft*. In: Eurostat, Allgemeine und Regionalstatistiken, Methodologies and Working Papers. Luxemburg 2008.

Gürünlü Alma, Özlem. *Comparison of Robust Regression Methods in Linear Regression*. In: International Journal of Contemporary Mathematical Sciences. 2011. Jahrgang 6, Nr. 9, Seite 409 ff.

Jung, Sandra/Käuser, Stefanie. [Herausforderungen und Potenziale der Einzeldatenverknüpfung in der Unternehmensstatistik](#). In: WISTA Wirtschaft und Statistik. Ausgabe 2/2016, Seite 95 ff.

Lorenz, Robin/Opfermann, Rainer. [Verwaltungsdaten in der Unternehmensstatistik](#). In: WISTA Wirtschaft und Statistik. Ausgabe 1/2017, Seite 49 ff.

Rousseeuw, Peter J. *Least Median of Squares Regression*. In: Journal of the American Statistical Association. 1984. Jahrgang 79, Nr. 388, Seite 871 ff.

Statistisches Bundesamt. *Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008)*. 2008. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Fachserie 4 Reihe 6.1. Produzierendes Gewerbe. Beschäftigung, Umsatz, Investitionen und Kostenstruktur der Unternehmen in der Energieversorgung, Wasserversorgung, Abwasser und Abfallentsorgung, Beseitigung von Umweltverschmutzungen*. 2018.

RECHTSGRUNDLAGEN

Empfehlung der Kommission (2003/361/EG) vom 6. Mai 2003 betreffend die Definition der Kleinstunternehmen sowie der kleinen und mittleren Unternehmen (Amtsblatt der EU Nr. L 124, Seite 36).

Gesetz ber Kostenstrukturstatistik (KoStrukStatG) vom 12. Mai 1959 in der im Bundesgesetzblatt Teil III, Gliederungsnummer 708-3, verffentlichten bereinigten Fassung, das zuletzt durch Artikel 4 des Gesetzes vom 21. Juli 2016 (BGBl. I Seite 1768) geandert worden ist.

Gesetz fr den Ausbau erneuerbarer Energien (Erneuerbare-Energien-Gesetz – EEG 2017) vom 21. Juli 2014 (BGBl. I Seite 1066), das zuletzt durch Artikel 5 des Gesetzes vom 13. Mai 2019 (BGBl. I Seite 706) geandert worden ist.

Gesetz ber die Statistik im Produzierenden Gewerbe (ProdGewStatG) in der Fassung der Bekanntmachung vom 21. Marz 2002 (BGBl. I Seite 1181), das zuletzt durch Artikel 271 der Verordnung vom 31. August 2015 (BGBl. I Seite 1474) geandert worden ist.

Gesetz ber die Verwendung von Verwaltungsdaten fr Zwecke der Wirtschaftsstatistiken (Verwaltungsdatenverwendungsgesetz – VwDVG) vom 4. November 2010 (BGBl. I Seite 1480), das durch Artikel 1 des Gesetzes vom 18. Dezember 2018 (BGBl. I Seite 2637) geandert worden ist.

Gesetz zur Neuordnung der Statistik im Produzierenden Gewerbe und zur anderung des Gesetzes ber Kostenstrukturstatistik vom 21. Marz 2002 (BGBl. I Seite 1181), das zuletzt durch Artikel 271 der Verordnung vom 31. August 2015 (BGBl. I Seite 1474) geandert worden ist.

Richtlinie 96/92/EG des Europaischen Parlaments und des Rates betreffend gemeinsame Vorschriften fr den Elektrizitatsbinnenmarkt vom 19. Dezember 1996 (Amtsblatt der EG Nr. L 027, Seite 20).

Zweites Gesetz zur Entlastung insbesondere der mittelstandischen Wirtschaft von Brokratie (Zweites Brokratieentlastungsgesetz) vom 30. Juni 2017 (BGBl. I Seite 2143).

Jana Emmenegger

studierte „Internationale Wirtschaft und Entwicklung“ an der Universität Bayreuth und „Angewandte Statistik“ an der Georg-August-Universität Göttingen. Seit 2018 ist sie wissenschaftliche Mitarbeiterin im Referat „Methoden der Datenanalyse“ des Statistischen Bundesamtes und promoviert dort – in Kooperation mit der Universität Trier – zum Thema Mikrosimulation.

Tobias Godec

ist Werkstudent im Referat „Methoden der Datenanalyse“ des Statistischen Bundesamtes und arbeitet ebenfalls zum Thema Mikrosimulation. Er schloss 2018 seinen Bachelor in „Soziologie und Volkswirtschaftslehre“ ab und studiert aktuell im Master „Wirtschafts- und Finanzsoziologie“ an der Goethe-Universität Frankfurt am Main.

Prof. Dr. Markus Zwick

ist Volkswirt und hat an der Leuphana Universität in Lüneburg promoviert. Im Statistischen Bundesamt arbeitete er zunächst im Bereich der Steuerstatistiken, von 2002 bis 2012 leitete er das Forschungsdatenzentrum. Nach vier Jahren als Nationaler Experte bei Eurostat übernahm er von Mitte 2016 bis 2017 die Leitung des Referats „Wissenschaftskooperation, Mikrosimulation, Neue digitale Daten“, danach die des Referats „Forschungsstrategie und -kooperationen, Wissenschaftskontakte“. Seit 2010 lehrt er als Honorarprofessor an der Goethe-Universität Frankfurt am Main.

SIMULATIONSPROJEKT ZU DIENSTUNFÄLLEN VON BEAMTINNEN UND BEAMTEN

Eine Mikrosimulation für das Bundesministerium
für Arbeit und Soziales

Jana Emmenegger, Tobias Godec, Prof. Dr. Markus Zwick

↳ **Schlüsselwörter:** Mikrosimulation – Arbeitsunfälle – synthetische Daten

ZUSAMMENFASSUNG

Dieser Artikel gibt einen Überblick darüber, welche statistischen Simulationsmodelle entwickelt und angewendet werden, um die Datenlieferverpflichtung Deutschlands zur Europäischen Statistik über Arbeitsunfälle zu erfüllen. Ab dem Berichtsjahr 2014 sind Daten über Dienstunfälle von Beamtinnen und Beamten zu übermitteln. Da solche Daten für den genannten Berichtskreis nicht vorlagen, sollten mithilfe von verfügbaren Daten und Simulationsmodellen die geforderten fallbezogenen Variablen für die Jahre 2014 bis 2016 als Einzeldatensätze abgebildet werden. So konnten die Anzahl an Dienstunfällen von Beamtinnen und Beamten und weitere Unfallattribute in einem Projekt des Statistischen Bundesamtes für das Bundesministerium für Arbeit und Soziales simuliert werden.

↳ **Keywords:** *microsimulation – work accidents – synthetic data*

ABSTRACT

This article gives an overview of the development and application of statistical simulation models in the context of Germany's obligation to transmit data for the European Statistics on Accidents at Work (ESAW). Data on work accidents of public officials have to be supplied as from reference year 2014. As such data were not available for public officials, the required case-related variables for the years 2014 to 2016 were compiled as individual data sets using available data and simulation models. In this way, the number of work accidents of public officials and additional accident attributes were simulated in a project of the Federal Statistical Office and the Federal Ministry of Labour and Social Affairs.

1

Einleitung

Gemäß der Verordnung (EU) Nr. 349/2011¹ bezüglich der Europäischen Statistik über Arbeitsunfälle (ESAW) hat Deutschland ab dem Berichtsjahr 2014 unter anderem Daten über Dienstunfälle von Beamtinnen und Beamten (Kommunal-, Landes- und Bundesebene) an das Statistische Amt der Europäischen Union (Eurostat) zu übermitteln. Die für den benannten Berichtskreis geforderten Daten liegen für die Jahre 2014 bis 2016 allerdings nicht vor. Ab dem Berichtsjahr 2017 können – vom Kalenderjahr 2019 an – die Lieferverpflichtungen durch eine Umstellung der Erfassungsmethodik erfüllt werden. Bei Nichtlieferung der Daten für Beamtinnen und Beamte für die Jahre 2014 bis 2016 an Eurostat drohte Deutschland ein Vertragsverletzungsverfahren.

Seit der Novellierung des Bundesstatistikgesetzes im Sommer 2016 ist es unter anderem die Aufgabe des Statistischen Bundesamtes, für oberste Bundesbehörden auf deren Anforderung *«Zusatzaufbereitungen für Bundeszwecke, einschließlich der Entwicklung und der Anwendung von Mikrosimulationsmodellen sowie mikroökonomischer Analysen durchzuführen»* (§ 3 Absatz 1 Nr. 6 Bundesstatistikgesetz). Zur Operationalisierung der neuen Aufgabe ist der Aufbau eines Methodenzentrums zur Mikrosimulation und -analyse geplant. Das Institut für Forschung und Entwicklung in der Bundesstatistik des Statistischen Bundesamtes ist daher verstärkt im Bereich der Mikrosimulation aktiv. Mitte 2017 hat es dem für die Übermittlung der Daten an Eurostat zuständigen Bundesministerium für Arbeit und Soziales Unterstützung für die Lieferung der fehlenden Daten angeboten: Dazu sollten im Rahmen eines Projekts die Daten zu Dienstunfällen von Beamtinnen und Beamten für die Jahre 2014 bis 2016 mittels eines statistischen Simulationsmodells geschätzt werden. Das hier vorgestellte Projekt ist eine erste Umsetzung der neuen Rechtsgrundlage und stellte eine Möglichkeit dar, Kompetenzen zum Thema Mikrosimulation auf- und auszubauen.

1 Verordnung (EU) Nr. 349/2011 der Kommission vom 11. April 2011 zur Durchführung der Verordnung (EG) Nr. 1338/2008 des Europäischen Parlaments und des Rates zu Gemeinschaftsstatistiken über öffentliche Gesundheit und über Gesundheitsschutz und Sicherheit am Arbeitsplatz betreffend Statistiken über Arbeitsunfälle (Amtsblatt der EU Nr. L 97, Seite 3).

Zur Umsetzung der konkreten Arbeiten schlossen das Bundesministerium für Arbeit und Soziales und das Statistische Bundesamt im Dezember 2017 eine Verwaltungsvereinbarung. Hiernach sind gemäß den Lieferverpflichtungen der EG-Verordnung Nr. 1338/2008 keine statistischen Tabellen, sondern Einzeldatensätze zu den Dienstunfällen zu erstellen. Der Konzeptentwurf wurde im Februar 2018 mit Eurostat abgestimmt, die Simulationsrechnungen im Dezember 2018 übermittelt.

2

Aufgabenstellung

Für die Europäische Statistik über Arbeitsunfälle hat Eurostat gemeinsam mit der Generaldirektion „Beschäftigung und Soziales“ der Europäischen Kommission eine zusammenfassende Methodik vorgelegt und publiziert (im weiteren ESAW-Methodik genannt). Sie beschreibt die zu liefernden Einzel- sowie Metadaten detailliert.² Diese ESAW-Methodik ist Grundlage der im Folgenden skizzierten Arbeiten.

Ziel des Projekts war es, mit den vorhandenen und verfügbaren Daten die geforderten fallbezogenen Variablen zu Dienstunfällen von Beamtinnen und Beamten für die Jahre 2014 bis 2016 bestmöglich im Rahmen der Zielvorgabe der ESAW-Methodik abzubilden.

3

Datengrundlage

Die nachfolgend beschriebenen Datenbestände wurden genutzt, um die simulierten Einzeldatensätze zu den Dienstunfällen von Beamtinnen und Beamten zu erstellen. Für die Jahre 2012 bis 2016 wurden dem Statistischen Bundesamt durch die Deutsche Gesetzliche Unfallversicherung die Daten der Europäischen Statistik über Arbeitsunfälle von Angestellten übermittelt. Des Weiteren lagen für den gleichen Zeitraum Einzeldaten der Personalstandstatistik sowie des Mikrozensus einschließlich des Ad-hoc-Moduls „Arbeitsunfälle“ der Arbeitskräfteerhebung 2013 vor.

2 <http://ec.europa.eu/eurostat>

3.1 Personalstandstatistik

Die Personalstandstatistik des öffentlichen Dienstes ermittelt jährlich zum Stichtag 30. Juni Daten über die Struktur des Personals im öffentlichen Dienst. Hierzu zählt das Personal aller Kernhaushalte³, Sonderrechnungen⁴ und der Einrichtungen in öffentlich-rechtlicher Rechtsform. Ihre Ergebnisse werden mit denen der Personalstandstatistik der öffentlich bestimmten Einrichtungen und Unternehmen in privater Rechtsform zu einem Gesamtbild der Beschäftigungssituation der öffentlichen Arbeitgeber ergänzt (Statistisches Bundesamt, 2018).

Der Datensatz der Personalstandstatistik diene als Grundlage für das Schätzmodell. Dabei handelt es sich um einen verdichteten Einzeldatensatz, das heißt jede Zeile des Datensatzes enthält Personalmerkmale und eine Fallzählervariable. Hierbei werden Fälle zusammengefasst, die identische Personalmerkmale (zum Beispiel Geschlecht, Geburtsjahr) aufweisen. Aus dem Material der Personalstandstatistik lassen sich Kennzahlen für die Grundgesamtheit der deutschen Bundes-, Landes- und Kommunalbeamtinnen und -beamten ableiten.

Die verwendete Personalstandstatistik umfasst alle in Deutschland Beschäftigten, die in einem unmittelbaren Dienst- oder Arbeitsvertragsverhältnis zu den auskunftspflichtigen Dienststellen stehen und in der Regel Gehalt oder Entgelt aus den Haushaltsmitteln dieser Stellen beziehen. Es werden unter anderem folgende Erhebungsmerkmale erfasst:

1. Alter, Geburtsmonat und -jahr
2. Geschlecht
3. Art, Umfang und Dauer des Dienst- oder Arbeitsvertragsverhältnisses
4. Dienst- oder Arbeitsort
5. Staatlicher und kommunaler Aufgabenbereich

Diese Merkmale lieferten relevante Informationen für das Simulationsmodell. Für das zu simulierende Merkmal des Wirtschaftszweigs lagen die Daten nicht in der entsprechenden Gliederung vor. Daher wurden die Anga-

³ Bund, Länder, Gemeinden/Gemeindeverbände, Sozialversicherung.

⁴ Rechtlich unselbstständige Einheiten in öffentlich-rechtlicher Rechtsform (zum Beispiel Bundes- und Landesbetriebe, kommunale Eigenbetriebe, Sondervermögen).

ben zum staatlichen und kommunalen Aufgabenbereich der Beamtinnen und Beamten nach der Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), in enger Zusammenarbeit mit der Fachabteilung umcodiert.

Die Zahl der Beamtinnen und Beamten im öffentlichen Dienst belief sich zum Stichtag 30. Juni 2016 auf etwa 1,67 Millionen; davon waren rund 180 000 im Bundesbereich, rund 1,3 Millionen im Landesbereich, 186 000 im kommunalen Bereich sowie rund 31 000 bei den Sozialversicherungsträgern einschließlich der Bundesagentur für Arbeit beschäftigt. Die Zahl der Arbeitnehmerinnen und Arbeitnehmer im öffentlichen Dienst betrug zum selben Stichtag insgesamt rund 2,8 Millionen: rund 146 000 im Bundesbereich, 1,09 Millionen im Landesbereich, 1,28 Millionen im Kommunalbereich und 340 000 bei den Sozialversicherungsträgern einschließlich der Bundesagentur für Arbeit.

3.2 Europäische Statistik über Arbeitsunfälle

Unternehmen haben jeden Unfall in ihrem Unternehmen anzuzeigen, wenn Versicherte getötet oder so verletzt sind, dass sie für vier oder mehr Tage arbeitsunfähig werden [meldepflichtiger Unfall nach §193 Absatz 1 Siebtes Buch Sozialgesetzbuch (SGB VII)]. Als Unfallereignis zählen alle Arbeitsunfälle im engeren Sinne (§8 Absatz 1 SGB VII). Wegeunfälle (§8 Absatz 2 SGB VII), also Unfälle, die sich auf dem Weg nach oder von dem Ort einer versicherten Tätigkeit ereignen, sind zwar versicherungsrechtlich den Arbeitsunfällen gleichgestellt. Sie werden jedoch von der Europäischen Statistik über Arbeitsunfälle nicht erfasst.

Bei den ESAW-Daten handelt es sich um Einzeldatensätze für Angestellte für die Jahre 2012 bis 2016, die 25 gemäß ESAW-Methodik definierte, unfallbezogene Variablen umfassen. Eine Liste der Variablen enthält die ESAW-Methodik (Europäische Kommission, 2013). Bei den ESAW-Daten handelt es sich nicht um die Gesamtheit der an die Deutsche Gesetzliche Unfallversicherung übermittelten Unfallmeldungen, sondern lediglich um eine Totalerhebung aller tödlichen Unfälle, ergänzt um eine Stichprobe anderer meldepflichtiger Arbeitsunfälle. Grund dafür ist, dass die Deutsche Gesetzliche Unfallversicherung für die Kodierung des Unfallhergangs gemäß der ESAW-Methodik mithilfe einer Zufalls-

lösung (Tag des Geburtstags) einzelne Fälle auswählt. Die vorliegenden Datensätze enthalten entsprechende Gewichte zur Hochrechnung auf die Grundgesamtheit der meldepflichtigen Arbeitsunfälle.

Die aggregierten Unfallzahlen werden von der Deutschen Gesetzlichen Unfallversicherung an das Bundesministerium für Arbeit und Soziales gemeldet. Für das Jahr 2016 umfasst die ESAW-Stichprobe 8,7% der hochgerechnet 63 154 nicht tödlichen Dienstunfälle und alle 23 tödlichen Unfälle, die den Unfallversicherungsträgern des öffentlichen Dienstes gemeldet wurden.

3.3 Mikrozensus

Bereits seit 1957 – in den neuen Ländern und Berlin-Ost seit 1991 – liefert der Mikrozensus jährlich statistische Informationen in tiefer fachlicher und regionaler Gliederung. Erhebungsmerkmale des Mikrozensus sind die Bevölkerungsstruktur, die wirtschaftliche und soziale Lage der Bevölkerung und der Familien, Lebensgemeinschaften und Haushalte, die Erwerbstätigkeit, Arbeitssuche, Aus- und Weiterbildung, Wohnverhältnisse und Gesundheit (Statistisches Bundesamt, 2017). Die Einzeldaten des Mikrozensus 2013 bis 2016 wurden für den Projektkontext genutzt. Explizit wurden folgende Erhebungsmerkmale für das Simulationsmodell herangezogen:

1. Art der Beschäftigung (Beamtinnen/Beamte, Arbeitnehmerinnen und Arbeitnehmer), Tätigkeit im öffentlichen Dienst
2. Berufsbezeichnung
3. Alter, Geburtsmonat und -jahr
4. Geschlecht
5. Art, Umfang und Dauer des Dienst- oder Arbeitsvertragsverhältnisses
6. Dienst- oder Arbeitsort
7. Aufgabenbereich

3.4 Ad-hoc-Modul „Arbeitsunfälle“ der europäischen Arbeitskräfteerhebung 2013

In den Mikrozensus integriert ist die Arbeitskräfteerhebung der Europäischen Union (EU) mit jährlichen Ad-hoc-Modulen. Diese dienen dazu, detaillierte Informationen zu sozialpolitisch und ökonomisch relevanten Themen der EU zu sammeln, welche nicht Teil des Standardfrageprogramms der Arbeitskräfteerhebung sind. Im Jahr 2013 wurde eine Zusatzerhebung zu Arbeitsunfällen und arbeitsbedingten Gesundheitsproblemen durchgeführt. Der Auswahlsatz betrug 0,1% der Bevölkerung, also 10% der Mikrozensusstichprobe (Statistisches Bundesamt, 2017).

Im Ad-hoc-Modul der Arbeitskräfteerhebung für das Jahr 2013 sind hochgerechnet insgesamt etwa 680 000 Arbeitsunfälle von Arbeitnehmerinnen und Arbeitnehmern und 33 630 Dienstunfälle von Beamtinnen und Beamten erfasst. Um die Vergleichbarkeit mit dem ESAW-Datensatz herzustellen, wurden Wegeunfälle, die sich auf dem Weg von oder zur Arbeitsstelle ereigneten, sowie Dienstunfälle, die eine Arbeitsunfähigkeit von weniger als vier Tagen verursachten, ausgeschlossen.

Insgesamt entfielen ungefähr zwei Drittel der Arbeitsunfälle auf Männer und ein Drittel auf Frauen. Dabei ist zu beachten, dass das geschlechtsspezifische Unfallrisiko mit der Verteilung der Geschlechter über die Wirtschaftszweige korreliert. Die im Ad-hoc-Modul der Arbeitskräfteerhebung 2013 erfassten Dienstunfälle von Beamtinnen und Beamten ereigneten sich ausschließlich in den Wirtschaftszweigen „öffentliche Verwaltung, Sozialversicherung“ (NACE⁵-Rev.-2-Sektor O) sowie „Erziehung und Unterricht“ (NACE-Rev.-2-Sektor P).

4

Analysekonzept

Auf Grundlage des zuvor beschriebenen Datenbestands wurde gemeinsam mit dem Mikrosimulationszentrum der Universität Trier ein statistisches Simulations-

⁵ NACE = Nomenclature statistique des activités économiques dans la Communauté européenne – Statistische Systematik der Wirtschaftszweige in der Europäischen Gemeinschaft.

modell zur Erstellung von synthetischen Einzeldaten-sätzen der Dienstunfälle von Beamtinnen und Beamten konzipiert. Das Modell basiert auf der Schichtung der Datensätze aus der Europäischen Statistik über Arbeitsunfälle und der Personalstandstatistik in möglichst homogene Gruppen entsprechend des Geschlechts, Wirtschaftszweigs, Alters und der Region. Anhand der beobachteten Unfallhäufigkeiten im Ad-hoc-Modul der Arbeitskräfteerhebung wurden modellgestützt Unfallwahrscheinlichkeiten berechnet. Dabei wurde eine Indikatorvariable aufgenommen, die den Wert 1 annimmt, wenn die Person verbeamtet ist, ansonsten 0. Mit diesem Ansatz ließ sich feststellen, ob es einen signifikanten Unterschied zwischen der Unfallwahrscheinlichkeit von Beamtinnen und Beamten sowie Arbeitnehmerinnen und Arbeitnehmern des öffentlichen Dienstes gibt. Auf Basis der geschätzten Unfallwahrscheinlichkeit wurden zufällig Beamtinnen und Beamte aus der jeweiligen Schicht der Personalstandstatistik gezogen und die fehlenden Unfallattribute ergänzend hinzugeschätzt. Durch die tiefgliedrige Schichtung der Personalstandstatistik wurde gewährleistet, dass die Heterogenität der Beamtinnen und Beamten sowie der Arbeitnehmerinnen und Arbeitnehmer adäquat in der Stichprobe widerspiegelt wird.

4.1 Schichtung der verschiedenen Datensätze

Mit den Einzeldaten aus der Personalstandstatistik ist die Grundgesamtheit der Arbeitnehmerinnen und Arbeitnehmer sowie der Beamtinnen und Beamten im öffentlichen Dienst bekannt. Bei der Personalstandstatistik handelt es sich um eine Vollerhebung administrativer Daten, die keinen Stichprobenfehler aufweisen. Daher wurde im Folgenden angenommen, dass die Personalstandstatistik für dieses Projekt die wahre Grundgesamtheit darstellt. Auf Basis der vorliegenden Merkmale in der Personalstandstatistik wurden Unterschiede der Gruppen von Arbeitnehmerinnen und Arbeitnehmern sowie Beamtinnen und Beamten im öffentlichen Dienst herausgearbeitet. Insbesondere wurde Heterogenität in der Altersstruktur, der regionalen Verteilung, dem Geschlechterverhältnis sowie der Verteilung auf die Wirtschaftszweige festgestellt. Zur Berücksichtigung dieser Unterschiedlichkeit im Simulationsmodell wurden sowohl der Datensatz aus der Personalstandstatistik

als auch der ESAW-Datensatz zunächst entsprechend dieser vier Variablen geschichtet.

Folgende Variablendefinitionen lagen zugrunde: Das Geschlecht hat zwei Ausprägungen $j = (1, 2)$, wobei 1 den männlichen Beamten und 2 die weibliche Beamtin kennzeichnet. Das Alter wurde in fünf Gruppen $l = (1, \dots, 5)$ zusammengefasst, die jeweils alle Personen enthalten, die (1) unter 26 Jahre, (2) von 26 bis unter 36 Jahre, (3) von 36 bis unter 46 Jahre, (4) von 46 bis unter 56 Jahre, oder (5) 56 Jahre und älter sind. Die regionale Gruppierung entspricht den Bundesländern $r = (1, \dots, 16)$. Eine tiefere regionale Gliederung liegt im ESAW-Datensatz nicht vor.

Die Wirtschaftssektoren $k \in (1, \dots, 10)$ wurden gruppiert. Nach Anhang II der Verordnung (EU) Nr. 349/2011 sind bestimmte Berufe und Wirtschaftszweige, deren Daten der Vertraulichkeit unterliegen, von der Pflicht zur Übermittlung der Daten an Eurostat ausgenommen. Hierunter fallen die Bereiche Verteidigung, Rechtspflege/Justiz und Öffentliche Sicherheit und Ordnung (Bundespolizei und Zoll). Unter Berücksichtigung dieser Regelung wurden 43 Wirtschaftszweige $k = (1, \dots, 43)$ nach NACE-Rev.-2-Klassifikation identifiziert, in denen Beamtinnen und Beamte tätig sind. Diese Wirtschaftszweige wurden in zehn Wirtschaftssektoren zusammengefasst, um zu gewährleisten, dass in jedem Sektor eine ausreichende

Übersicht 1

Gruppierung der Wirtschaftszweige

Wirtschafts-sektor	NACE-Rev.-2-Beschreibung
1	A: Land- und Forstwirtschaft, Fischerei C: Verarbeitendes Gewerbe/Herstellung von Waren F: Baugewerbe/Bau
2	D: Energieversorgung E: Wasserversorgung; Abwasser- und Abfallentsorgung und Beseitigung von Umweltverschmutzungen
3	H: Verkehr und Lagerei
4	I: Gastgewerbe/Beherbergung und Gastronomie R: Kunst, Unterhaltung und Erholung
5	J: Information und Kommunikation L: Grundstücks- und Wohnungswesen
6	K: Erbringung von Finanz- und Versicherungsdienstleistungen N: Erbringung von sonstigen wirtschaftlichen Dienstleistungen S: Erbringung von sonstigen Dienstleistungen
7	M: Erbringung von freiberuflichen, wissenschaftlichen und technischen Dienstleistungen
8	Q: Gesundheits- und Sozialwesen
9	O: Öffentliche Verwaltung; Sozialversicherung
10	P: Erziehung und Unterricht

Anzahl von Beamtinnen und Beamten enthalten ist. Eine konsistente Kodierung der Variablen über die verschiedenen Datenquellen hinweg wurde sichergestellt.

↳ [Übersicht 1](#)

Aus der Kombination von Geschlecht $j \in (1, 2)$, Wirtschaftssektor $k \in (1, \dots, 10)$, Altersgruppe $l \in (1, \dots, 5)$ und Bundesland $r \in (1, \dots, 16)$ ergaben sich $2 \cdot 10 \cdot 5 \cdot 6 = 1600$ mögliche Schichten. Durch Zusammenfassung von kleinen, mit weniger als 100 Personen besetzten Schichten wurden insgesamt 336 Schichten gebildet. ↳ [Übersicht 2](#)

Übersicht 2

Stratifizierungsplan

Geschlecht	Wirtschaftssektor (siehe Übersicht 1)	Region (Land)	Altersgruppe (in Jahren)	Schicht ID
Männer	1	alle	alle	1
	...	alle	alle	...
	8	alle	alle	8
	9	Land 01	unter 26	9
			26 bis unter 36	10
			36 bis unter 46	11
			46 bis unter 56	12
			56 und älter	13
	14 – 83	
	Land 16	wie Land 01	84 – 88	
10	wie Wirtschaftssektor 9	89 – 168		
Frauen	Aufbau wie oben			169 – 336

4.2 Schätzung des Unfallrisikos

Basierend auf dem Ad-hoc-Modul der Arbeitskräfteerhebung 2013 wurde eine modellbasierte Schätzung detaillierter Unfallwahrscheinlichkeiten unter Berücksichtigung demografischer und arbeitsbezogener Variablen X_{s_i} für alle Personen i der Schichten s durchgeführt. Mithilfe von Kreuzvalidierungsmethoden wurde aus verschiedenen Verfahren das gewichtete Logit-Modell ausgewählt. Es wurde angenommen, dass Unfälle U_{s_i} nach einer Bernoulli-Wahrscheinlichkeitsfunktion mit der Wahrscheinlichkeit $\hat{\pi}_{s_i}^{MC}$ den Wert 1 und mit der Wahrscheinlichkeit $1 - \hat{\pi}_{s_i}^{MC}$ den Wert 0 annehmen:

$$U_{s_i} \sim \text{Bernoulli}(U_{s_i} | \hat{\pi}_{s_i}^{MC})$$

Die Unfallwahrscheinlichkeiten von Beamtinnen und Beamten sowie Arbeitnehmerinnen und Arbeitnehmern des öffentlichen Dienstes wurden durch ein Logit-Modell geschätzt:

$$(1) \hat{\pi}_{s_i}^{MC} = P(U_{s_i} = 1 | X = X_{s_i}) = f_{\beta}(X_{s_i}) = \frac{1}{1 + e^{-X_{s_i}\beta}},$$

wobei $\hat{\pi}_{s_i}^{MC}$ die geschätzte Unfallwahrscheinlichkeit ist, X_{s_i} ein Vektor von Kovariablen für Person i in Schicht s einschließlich Geschlecht, Altersgruppe, Region, Wirtschaftszweiggruppe und der Indikatorvariablen Verbeamtet. $\beta = (\beta_0, \dots, \beta_n)$ ist der Vektor der entsprechenden Koeffizienten. Für weitere Details zur logistischen Regression siehe beispielsweise King/Zeng (2001).

4.3 Zufällige Ziehung der geschädigten Beamtinnen und Beamten

Nun konnte die erwartete Anzahl an Dienstunfällen von Beamtinnen und Beamten \hat{U}_s^B für jede Schicht der Personalstandstatistik wie folgt ermittelt werden:

$$(2) \hat{U}_{s,t}^B = \hat{\pi}_s^{MC} \cdot N_{s,t}^B$$

mit

$$(3) \hat{\pi}_s^{MC} = \sum_{i=1}^{N_s} w_i \cdot \hat{\pi}_{s_i}^{MC}$$

Dabei entspricht $N_{s,t}^B$ der Anzahl der Beamtinnen und Beamten in der jeweiligen Schicht $s \in (1, \dots, 336)$ der Personalstandstatistik für jedes Jahr $t \in (2014, 2015, 2016)$. Die geschätzte Unfallwahrscheinlichkeit $\pi_{s_i}^{MC}$ in Schicht $s \in (1, \dots, 336)$ ist ein gewichteter Mittelwert der individuellen Unfallwahrscheinlichkeiten $\hat{\pi}_{s_i}^{MC}$, wobei w_i der im Ad-hoc-Modul der Arbeitskräfteerhebung bereitgestellte Hochrechnungsfaktor für jedes Individuum i ist.

Im nächsten Schritt wurde die geschätzte Anzahl der verunfallten Beamtinnen und Beamten, $\hat{U}_{s,t}^B$, zufällig ausgewählt. Dies erfolgte ohne Zurücklegen aus der relevanten Schicht $s \in (1, \dots, 336)$, der Statistik des Personals im öffentlichen Dienst für jedes Jahr $t \in (2014, 2015, 2016)$ unter Berücksichtigung der geschätzten Unfallwahrscheinlichkeit jeder einzelnen Person $\hat{\pi}_{s_i}^{MC}$. Alle identifizierten Personen wurden in einem separaten Datensatz gespeichert, der dann mit der ESAW-Statistik für die Arbeitnehmerinnen und Arbeitnehmer des jeweiligen Jahres zusammengeführt wurde.

Die Varianz der gesamten Zahl der geschätzten Dienstunfälle $\hat{U}_{s,t}^B$ ist gegeben durch:

$$(4) \quad \widehat{Var}(\hat{U}_{s,t}^B) = Var(\hat{\pi}_s^{MC} \cdot N_{s,t}^B) = \sum_{s=1}^S \frac{N_s (N_s - n_s)}{n_s} \cdot \hat{\sigma}_s^2$$

Dabei bezeichnen S , N_s , n_s die Nummer der Schichten, die Populationsgröße der Schicht s und die Stichprobengröße innerhalb der Schicht $s \in (1, \dots, 336)$. Die Stichprobenvarianz der endlichen Population $\hat{\sigma}_s^2$ in der Schicht s ist gegeben als:

$$(5) \quad \hat{\sigma}_s^2 = \frac{1}{n_s - 1} \sum_{i=1}^{n_s} (X_{s_i} - \bar{X}_s)^2,$$

wobei $\bar{X}_s = \frac{1}{n_s} \sum_{i=1}^{n_s} X_{s_i}$ und alle übrigen Variablen wie oben definiert sind (Cao und andere, 2013).

4.4 Simulation der unfallspezifischen Merkmale

Der resultierende Datensatz liefert Informationen über den Wirtschaftszweig, das Alter, den Dienstort und das Geschlecht jeder Beamtin und jedes Beamten. Der Dienstort des identifizierten Unfallopfers wurde als Proxy für den Unfallort herangezogen. Um die übrigen nach der ESAW-Methodik definierten unfallspezifischen Variablen zu ergänzen, wurde eine Vielzahl verschiedener Methoden auf ihre Eignung überprüft und – soweit sinnvoll – angewendet. Das wichtigste Entscheidungskriterium für die Wahl der am besten geeigneten Methode für jede simulierte Variable war die Berücksichtigung des jeweiligen Datentyps. So ist beispielsweise die logistische Regressionsimputation für diskrete Auswahlvariablen geeignet. Darüber hinaus wurde die Methode auf der Grundlage der verfügbaren relevanten Informationen ausgewählt. Der Simulationsprozess einiger ausgewählter Variablen wird im Folgenden beschrieben.

Um beispielsweise das [Datum des Dienstunfalls](#) für alle Beamtinnen und Beamten zu simulieren, wurde die Annahme, dass das Auftreten von Dienstunfällen gleichmäßig über das Jahr verteilt ist, mithilfe eines χ^2 -Tests geprüft. Die Testergebnisse zeigten, dass innerhalb aller zweistelligen Wirtschaftszweige die Nullhypothese der Gleichverteilung der Dienstunfälle über die Zeit nicht verworfen werden kann. Daher wurde in jedem Wirtschafts-

zweig nach dem Zufallsprinzip ein Datum für jeden Dienstunfall einer Beamtin oder eines Beamten ausgewählt.

Zur Simulation der [Berufsgruppe des Geschädigten](#) wurde eine Cold-Deck-Imputation auf der Grundlage der Berufsgruppen von Beamtinnen und Beamten im Mikrozensus durchgeführt. Dagegen kam für die Simulation der [Art der Verletzung](#) und des [betroffenen Körperteils](#) eine Hot-Deck-Imputation basierend auf den ESAW-Daten zum Einsatz (Andridge/Little, 2010; van Buuren, 2018, hier: Kapitel 3.5). Innerhalb der Schichten wurde für jede fehlende Merkmalsausprägung ein beobachteter Wert zufällig gezogen. Hierbei wurde eine Gewichtung entsprechend der Anzahl der Beamtinnen und Beamten in der jeweiligen Schicht berücksichtigt.

Die vier Variablen [Arbeitsumgebung](#), [Kontakt – Art der Verletzung](#), [Abweichung](#) und [Gegenstand der Abweichung](#) wurden auf Basis von Klassifikations- und Regressionsbäumen (CART) simuliert. CART ist ein maschinelles Lernverfahren, das bei jedem Schritt eine Aufteilung der Beobachtungen in zwei Teilmengen vornimmt. Dabei sucht es automatisch Prädiktoren und Cut-Points, die die Stichprobe so teilen, dass möglichst homogene Gruppen entstehen. Jede Gruppe wurde immer weiter in jeweils zwei Untergruppen geteilt, sodass ein binärer Baum entstanden ist. Aus der Spendergruppe nach der letzten Teilung wurde zufällig ein Spender gezogen, dessen Wert imputiert wurde. Dieses Vorgehen benötigt keine Verteilungsannahmen und ist robust gegen Ausreißer. Es berücksichtigt die Abhängigkeiten zwischen Variablen und gewährleistet so plausible Einzeldatensätze, insbesondere in dem hier vorliegenden Fall von mehreren fehlenden, miteinander korrelierten Variablen. Zudem ist die Vorgehensweise nicht subjektiv, sondern maschinell gesteuert (van Buuren, 2018, hier: Kapitel 3.5).

5

Ergebnisse der Simulation

Im Folgenden werden ausgewählte Auszüge aus den Ergebnissen des Simulationsmodells sowie aus den Vergleichen der beobachteten und simulierten Verteilungen vorgestellt.

5.1 Schätzung der Unfallwahrscheinlichkeit

↘ **Tabelle 1** zeigt die Ergebnisse der Schätzung der Unfallwahrscheinlichkeiten mittels eines gewichteten logistischen Regressionsmodells auf Basis des Ad-hoc-Moduls der Arbeitskräfteerhebung aus Gleichung (2) in Abschnitt 4.3.

Die erste Spalte von Tabelle 1 enthält die geschätzten Koeffizienten $\hat{\beta}$ der erklärenden Variablen aus dem Logit-Modell. Es ist deutlich zu erkennen, dass die Unfallwahrscheinlichkeit bei Frauen signifikant geringer ist als bei Männern. Darüber hinaus sinkt die Wahrscheinlichkeit eines Arbeitsunfalls mit zunehmendem Alter, das heißt jüngere Mitarbeiterinnen und Mitarbeiter sind stärker gefährdet. Aus Zeile 4 ist ersichtlich, dass der Koeffizient der Indikatorvariable Beamte nicht signifikant ist. Das bedeutet, dass die Wahrscheinlichkeit eines Arbeitsunfalls unter sonst gleichen Bedingungen bei Beamtinnen und Beamten sowie Arbeitnehmerinnen und Arbeitnehmern nicht signifikant unterschiedlich ist.

Basierend auf den Ergebnissen der Regression wurden die Unfallwahrscheinlichkeiten für Beamtinnen und Beamte durch Anwendung der geschätzten Koeffizienten auf die Daten der Personalstandstatistik vorhergesagt. Anschließend wurden für alle 336 Schichten der Personalstandstatistik die Unfallwahrscheinlichkeiten und die erwartete Anzahl der Dienstunfälle berechnet, wie in den Gleichungen (3) und (4) aus Abschnitt 4.4 erläutert. Einen Auszug aus diesen Ergebnissen zeigt ↘ **Tabelle 2** auf Seite 52.

Für das Jahr 2014 wurden insgesamt 55 549 Dienstunfälle simuliert. Die geschätzte durchschnittliche Unfallwahrscheinlichkeit insgesamt beträgt 0,032 für die Referenzgruppe der 1,72 Millionen Beamtinnen und Beamten. Um die Unsicherheit der geschätzten Gesamt-

Tabelle 1
Gewichteter Logit-Output

	Abhängige Variable: Dienstunfall	
	Koeffizient	Standardfehler
Konstante	3,526***	0,203
Frauen	0,286***	0,075
Alter	0,008***	0,003
Beamte	0,091	0,154
Baden-Württemberg	0,279	0,349
Bayern	0,709***	0,137
Berlin	0,285	0,188
Brandenburg	0,600*	0,308
Bremen	0,208	0,257
Hamburg	0,348***	0,131
Hessen	0,286	0,241
Mecklenburg-Vorpommern	0,427**	0,197
Niedersachsen	0,142	0,237
Nordrhein-Westfalen	0,502	0,603
Rheinland-Pfalz	0,120	0,174
Sachsen	0,286	0,269
Sachsen-Anhalt	0,579*	0,311
Schleswig-Holstein	1,255***	0,116
Thüringen	0,038	0,309
NACE A, C, F	0,485***	0,143
NACE D, E	1,003*	0,537
NACE G	0,067	0,164
NACE H	0,499**	0,195
NACE I, R	0,475**	0,201
NACE J, L	0,024	0,236
NACE M	0,054	0,205
NACE O	0,180	0,183
NACE P	0,418**	0,176
NACE Q	0,419***	0,155

*, ** und *** geben statistisch signifikante Koeffizienten für $p = 0,10, 0,05$ und $0,01$ an. Die Referenzkategorie ist die Kategorie mit den meisten Beobachtungen, hier: Saarland und NACE K, S, N (Dienstleistungssektor).

summe der Dienstunfälle zu quantifizieren, wurde der entsprechende Stichprobenfehler unter Berücksichtigung des geschichteten Stichprobenverfahrens geschätzt. Die Korrektur für endliche Populationen und die Korrektur für Stichproben mit Ziehen ohne Zurücklegen waren nicht erforderlich, da die Zahl der Dienstunfälle relativ gering ist. Der geschätzte Standardfehler beträgt 230,93 für die Summe der Dienstunfälle von 55 549. Das 95-%-Konfidenzintervall für die Gesamtzahl der Dienstunfälle liegt somit bei 55 096 bis 56 002.

Tabelle 2

Ergebnisse nach Wirtschaftszweigen für 2014

Schicht id	Beschreibung	Männer			Frauen		
		Anzahl Beamte	Unfallwahrscheinlichkeit	Anzahl Dienstunfälle	Anzahl Beamtinnen	Unfallwahrscheinlichkeit	Anzahl Dienstunfälle
1	NACE A, C, F	6 784	0,044	301	1 011	0,036	36
2	NACE D, E	1 585	0,009	15	849	0,007	6
3	NACE H	32 384	0,042	1 373	4 394	0,033	143
4	NACE I, R	2 742	0,043	119	3 204	0,032	102
5	NACE J, L	1 638	0,028	45	1 085	0,021	23
6	NACE K, S, N	4 588	0,026	120	2 114	0,020	43
7	NACE M	1 968	0,025	49	846	0,020	16
8	NACE Q	8 949	0,034	304	4 391	0,030	130
9	Durchschnitt aller NACE-O-Schichten	587 005	0,033	19 289	348 981	0,025	8 746
10	Durchschnitt aller NACE-P-Schichten	243 513	0,042	9 989	465 683	0,032	14 700
Insgesamt		891 156	0,035	31 604	832 458	0,028	23 945

Insgesamt (Männer und Frauen) erwartet: Anzahl Beamtinnen und Beamte: 1,72 Millionen; Unfallwahrscheinlichkeit: 0,032; Anzahl Dienstunfälle: 55 549.

5.2 Simulierte gegenüber beobachtete Verteilungen

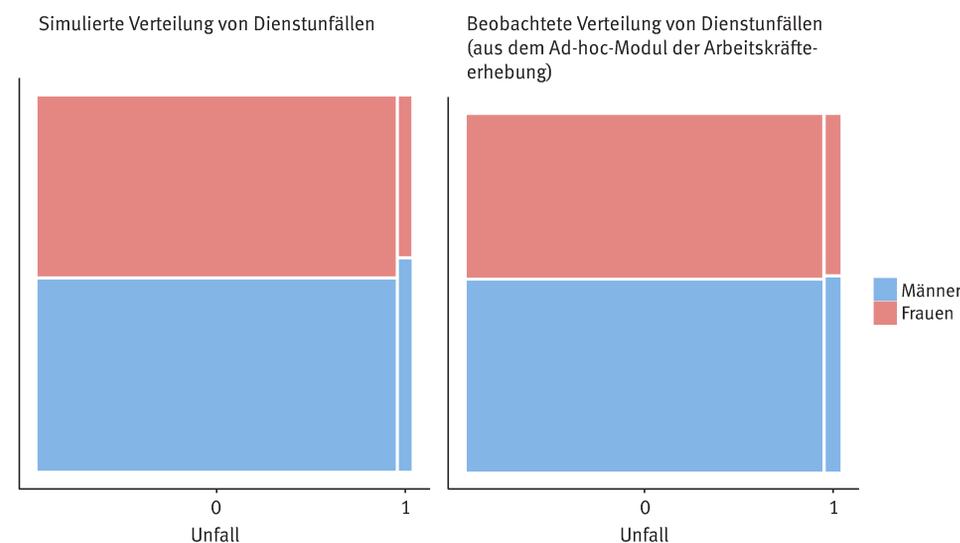
Zur visuellen Überprüfung der ermittelten Unfallzahlen wurden für alle Kombinationen der Variablen Alter, Geschlecht und Wirtschaftszweig der Beamtinnen und Beamten simulierte und beobachtete Verteilungen systematisch verglichen. Wenn die simulierte Verteilung

von der beobachteten Verteilung abweicht, wurden mögliche Gründe dafür herausgearbeitet.

Zunächst wurde die simulierte Verteilung der Dienstunfälle im öffentlichen Dienst nach Geschlecht untersucht, wie in [Grafik 1](#) dargestellt. Die linke Seite des Mosaik-Plots zeigt die simulierten Ergebnisse, die rechte Seite die beobachtete Verteilung der Dienstunfälle nach Geschlecht. Letztere wurde auf Grundlage der Daten

Grafik 1

Mosaik-Plot von Dienstunfällen von Beamtinnen und Beamten



Ausprägung 0 = kein Unfall; 1 = Unfall

2019 - 01 - 0529

aus dem Ad-hoc-Modul der Arbeitskräfteerhebung erstellt, das als einzige Datenquelle Informationen über die Dienstunfälle von Beamtinnen und Beamten enthält. Die Fläche der Kästen ist proportional zu den Zellhäufigkeiten der Kreuztabelle der Dienstunfälle nach Geschlecht. Die Breite der Kästen stellt den Prozentsatz der Arbeitsunfälle der Beamtinnen und Beamten dar. Während die simulierte Verteilung ergibt, dass 3,3% der Beamtinnen und Beamten einen Dienstunfall hatten, ist die beobachtete Verteilung

lung mit 2,9% sehr ähnlich. Darüber hinaus zeigen die Farben der Kästen die Verteilung der Dienstunfälle nach Geschlecht. 43,1% aller simulierten beziehungsweise 44,1% aller beobachteten Dienstunfälle sind Frauen widerfahren. Das Geschlechterverhältnis bei den Dienstunfällen und das Gesamtniveau der Dienstunfälle wurden daher in der Simulation recht gut berücksichtigt.

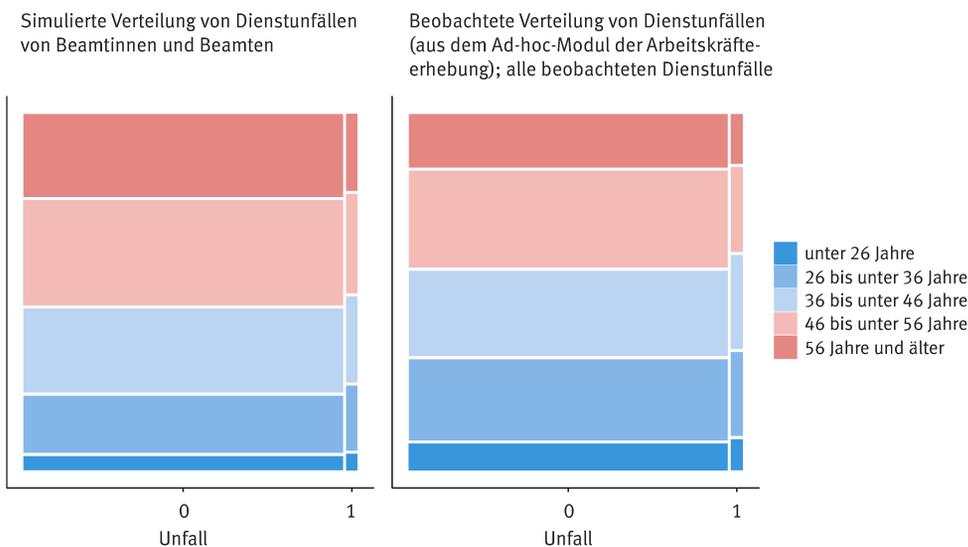
↳ Grafik 2 zeigt die simulierte Verteilung der Arbeitsunfälle nach Altersgruppen. Diese kann in Anlehnung an das erste Beispiel interpretiert werden. Zu beachten ist, dass in der Grafik 2 simulierte Dienstunfälle von Beamtinnen und Beamten auf der linken Seite und alle beobachteten Dienstunfälle (einschließlich Arbeitnehmerinnen und Arbeitnehmern) auf der rechten Seite einander gegenübergestellt wurden. Grund dafür ist, dass die Unfallvariable im Ad-hoc-Modul der Arbeitskräfteerhebung nicht alle Altersgruppen der Beamtinnen und Beamten abdeckt. Dadurch gibt es keine vollständigen Informationen über die beobachtete Verteilung der Dienstunfälle von Beamtinnen und Beamten nach Altersgruppen. Vergleicht man die Höhe der unterschiedlich farbigen Kästen nach Dienstunfällen, so lässt sich eine leichte Unterschätzung der Dienstunfälle bei jüngeren Beamtinnen und Beamten (blau dargestellt) und eine leichte Überschätzung bei Beamtinnen und Beamten höherer Altersgruppen (in Rottönen dargestellt) feststel-

len. Dies könnte daran liegen, dass die Altersstruktur der Beamtinnen und Beamten im Vergleich zu der der Arbeitnehmerinnen und Arbeitnehmer (im rechten Teil der Grafik) heterogen ist.

↳ Grafik 3 auf Seite 54 untersucht, ob die beobachtete Verteilung der Beamtinnen und Beamten aus der Personalstandstatistik nach Wirtschaftszweigen und Geschlecht in der Simulation beibehalten wurde. Es ist zu erkennen, dass die Wirtschaftszweige „Öffentliche Verwaltung und Sozialversicherung“ (NACE-Rev.-2-Sektor O) und „Erziehung und Unterricht“ (NACE-Rev.-2-Sektor P) etwa 95% aller Beamtinnen und Beamten in beiden Verteilungen abdecken. Durch die geschichtete Stichprobenmethode wurden auch alle anderen Wirtschaftszweige, in denen Beamtinnen und Beamte beobachtet wurden, in die Simulation miteinbezogen. In den simulierten Daten ist der Anteil der Männer geringfügig größer als das beobachtete Geschlechterverhältnis, da – wie in Grafik 1 dargestellt – Männer häufiger Dienstunfälle erleiden.

Grafik 2

Mosaik-Plot von Dienstunfällen von Beamtinnen/Beamten sowie Arbeitnehmerinnen/Arbeitnehmern nach Altersgruppen

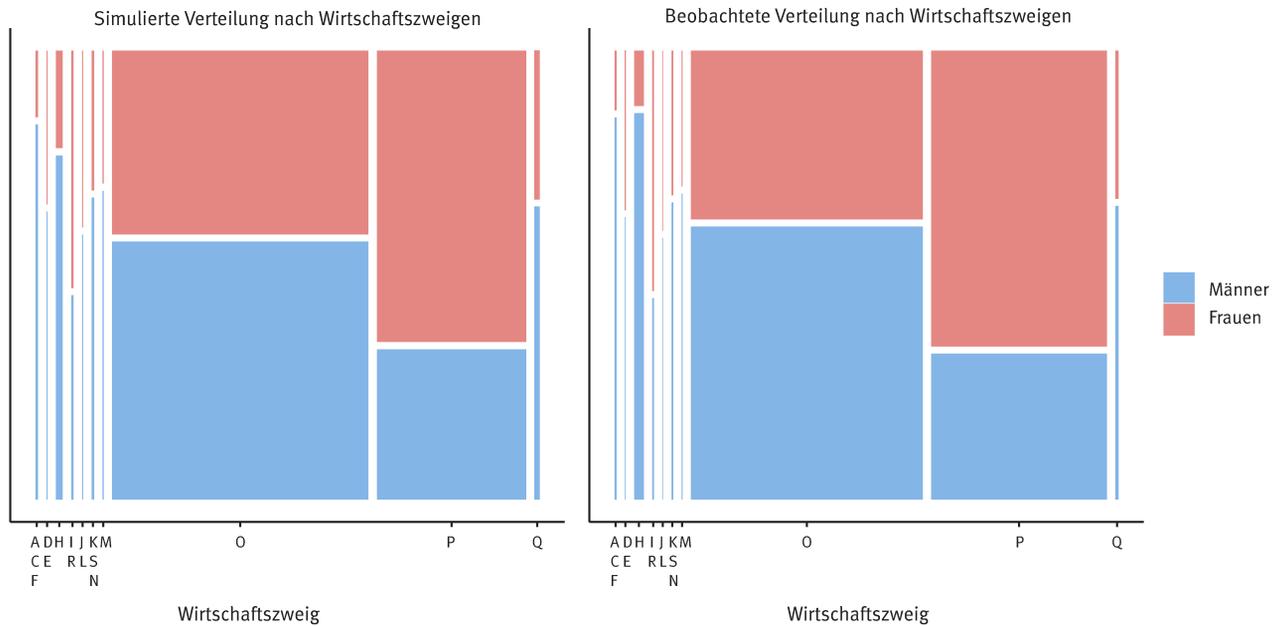


Ausprägung 0 = kein Unfall; 1 = Unfall

2019 - 01 - 0530

Grafik 3

Mosaik-Plot von Beamtinnen und Beamten nach Wirtschaftszweigen



2019 - 01 - 0531

6

Fazit und Ausblick

Zusammenfassend lässt sich sagen, dass die Daten zu Dienstunfällen von Beamtinnen und Beamten erfolgreich für die Jahre 2014 bis 2016 mittels eines statistischen Simulationsmodells geschätzt werden konnten. Die vorhandenen und verfügbaren Datengrundlagen wurden umfassend geprüft und vier relevante Datenquellen identifiziert, die für das Projekt herangezogen werden konnten. Basierend auf verschiedenen methodischen Ansätzen wurde ein Modell entwickelt und umgesetzt, um die geforderten fallbezogenen Variablen zu Dienstunfällen von Beamtinnen und Beamten für die Jahre 2014 bis 2016 bestmöglich im Rahmen der Zielvorgabe der ESAW-Methodik abzubilden. Abschließend wurden die simulierten und beobachteten Verteilungen aller Variablen verglichen und ein Stichprobenfehler berechnet, um die Qualität der Simulation zu beurteilen. Insgesamt stellte das Projekt eine gute Möglichkeit dar, Erfahrung im Bereich Mikrosimulation zu sammeln: Es wurden Daten simuliert, die anderenfalls für den betrachteten Berichtskreis der Beamtinnen und Beamten nicht vorgelegen hätten.

Zur weiteren Verbesserung des Simulationsmodells könnten multiple Imputationsmethoden angewendet werden, um zuverlässigere Schätzungen der Varianz der oben dargestellten simulierten Variablen zu erhalten. Als zusätzliche Qualitätsmessung könnten die für die Jahre 2014 bis 2016 simulierten Datensätze rückwirkend mit Unfalldaten verglichen werden, die von der Unfallversicherung Bund und Bahn sowie – soweit vorhanden – von den Unfallversicherungsträgern der Länder später geliefert wurden. Es ist vorgesehen, das Simulationsverfahren auch für 2017 anzuwenden, um die daraus resultierenden Daten mit den ersten gemeldeten Daten für Beamtinnen und Beamten zu vergleichen, sobald diese vorliegen. [u](#)

LITERATURVERZEICHNIS

Andridge, Rebecca R./Little, Roderick J. A. *A review of hot deck imputation for survey non-response*. In: International Statistical Review. Jahrgang 78. Ausgabe 1/2010, Seite 40 ff.

Cao, Ricardo/Vilar, José A./Vilar, Juan M./López, Ana K. *Sampling Error Estimation in Stratified Surveys*. In: Open Journal of Statistics. Jahrgang 3. Ausgabe 3/2013, Seite 200 ff. DOI: 10.4236/ojs.2013.33023

Europäische Kommission. *Europäische Statistik über Arbeitsunfälle*. 2013. [Zugriff am 26. Juni 2019]. Verfügbar unter: <https://ec.europa.eu/eurostat>

Statistisches Bundesamt. *Finanzen und Steuern – Personal des öffentlichen Dienstes 2017*. 2018. [Zugriff am 28. Juni 2019]. Verfügbar unter: www.destatis.de

Statistisches Bundesamt. *Qualitätsbericht – Mikrozensus 2016*. [Zugriff am 28. Juni 2019]. Verfügbar unter: www.destatis.de

Van Buuren, Stef. *Flexible Imputation of Missing Data*. 2018. In: Chapman and Hall/CRC. [Zugriff am 27. Juni 2019]. Verfügbar unter: <https://stefvanbuuren.name/fimd/>

RECHTSGRUNDLAGEN

Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I Seite 2394), das zuletzt durch Artikel 10 Absatz 5 des Gesetzes vom 30. Oktober 2017 (BGBl. I Seite 3618) geändert worden ist.

Siebttes Buch Sozialgesetzbuch – Gesetzliche Unfallversicherung – (Artikel 1 des Gesetzes vom 7. August 1996, BGBl. I Seite 1254), das zuletzt durch Artikel 9 des Gesetzes vom 6. Mai 2019 (BGBl. I Seite 646) geändert worden ist.

Verordnung (EU) Nr. 349/2011 der Kommission vom 11. April 2011 zur Durchführung der Verordnung (EG) Nr. 1338/2008 des Europäischen Parlaments und des Rates zu Gemeinschaftsstatistiken über öffentliche Gesundheit und über Gesundheitsschutz und Sicherheit am Arbeitsplatz betreffend Statistiken über Arbeitsunfälle (Amtsblatt der EU Nr. L 97, Seite 3).

Verordnung (EG) Nr. 1338/2008 des Europäischen Parlaments und des Rates vom 16. Dezember 2008 zu Gemeinschaftsstatistiken über öffentliche Gesundheit und über Gesundheitsschutz und Sicherheit am Arbeitsplatz (Amtsblatt der EU Nr. L 354, Seite 70).



Nelli Krüger

ist Diplom-Psychologin und Referentin im Referat „Veröffentlichungen, Öffentlichkeitsarbeit, Internationale Kontakte“ im Bereich Haushaltserhebungen des Statistischen Bundesamtes. Sie beschäftigt sich vor allem mit der Konzeption von Kommunikationsmaßnahmen zur Steigerung der Teilnahmebereitschaft und Akzeptanz von Haushaltserhebungen.

MASSNAHMEN ZUR STEIGERUNG DER TEILNAHMEBEREITSCHAFT AN HAUSHALTSERHEBUNGEN DER AMTLICHEN STATISTIK

Nelli Krüger

↳ **Schlüsselwörter:** Haushaltserhebungen – Teilnahmebereitschaft – Responserate – Einkommens- und Verbrauchsstichprobe – Incentives

ZUSAMMENFASSUNG

Haushaltserhebungen der amtlichen Statistik liefern unverzichtbare Daten für Politik, Wirtschaft, Wissenschaft und die breite Öffentlichkeit. Als Primärerhebungen spiegeln sie in einer einzigartigen Weise die wirtschaftliche und soziale Situation von Privathaushalten wider. Für eine hohe Datenqualität ist auch eine hohe Teilnehmerate bei den Erhebungen entscheidend. Jedoch ist die Bereitschaft privater Haushalte, an Erhebungen der amtlichen Statistik teilzunehmen, seit Jahren rückläufig beziehungsweise stagniert auf niedrigem Niveau. Im vorliegenden Aufsatz wird dargestellt, dass zur Steigerung der Teilnahmebereitschaft der Einsatz eines breiten Maßnahmenbündels erforderlich ist. Hierzu werden am Beispiel der Einkommens- und Verbrauchsstichprobe konkrete Maßnahmen vorgestellt.

↳ **Keywords:** household surveys – willingness to participate – response rate – sample survey of income and expenditure – incentives

ABSTRACT

Household surveys of official statistics provide indispensable data for policy-makers, businesses, science and the general public. Such primary surveys uniquely reflect the economic and social situation of households. High rates of participation in the surveys are crucial for good data quality. However, the willingness of households to participate in official statistical surveys has been declining or stagnating at a low level for years. This article shows that in order to increase the willingness to participate it is necessary to use a broad package of measures. For this purpose concrete measures are presented by using the example of the sample survey of income and expenditure.

1

Entwicklung der Teilnahmebereitschaft

Über sinkende Ausschöpfungsquoten bei Haushaltsbeziehungswise Personenbefragungen wird bereits seit mehreren Jahrzehnten berichtet. In einem Ländervergleich wurden Daten aus zehn amtlichen Erhebungen in 16 Ländern, die teilweise bis in die 1970er-Jahre zurückreichen, analysiert. Dabei wurde festgestellt, dass bei Zufallsstichproben die Responserate über die Zeit sukzessive gesunken ist (de Leeuw/de Heer, 2002). Die entscheidenden Faktoren, die die Responserate bestimmen, sind die Erreichbarkeit und die Verweigerungsrate von Befragungspersonen. Analysen für den Zeitraum 1972 bis 1998 zeigten, dass die Nichterreichbarkeitsrate jährlich um 0,2% und die Verweigerungsrate um 0,3% angestiegen ist. Die zunehmend schwierigere Erreichbarkeit der Befragungspersonen sowie deren zunehmende Verweigerungsrate führen somit zur stetig sinkenden Ausschöpfungsrate. Auch für den Zeitraum von 1998 bis 2015 wurden die international sinkenden Ausschöpfungsquoten bestätigt (Luiten und andere, 2016).

Der Mikrozensus ist die größte jährlich durchgeführte amtliche Haushaltserhebung in Deutschland. Die zufällig ausgewählten Haushalte sind gesetzlich zur Teilnahme verpflichtet. Dementsprechend ist die Teilnahmequote hoch. Für einzelne Fragen besteht jedoch keine Auskunftspflicht. Hier zeigt sich eine zum Teil hohe Ausfallquote, insbesondere bei den Haushalten, die den Fragebogen selbstständig ohne eine Interviewerin oder einen Interviewer ausfüllen. So liegt bei diesen Haushalten für Merkmale ohne Auskunftspflicht, wie das alle vier Jahre im Zusatzprogramm „Gesundheit“ erfasste Körpergewicht, die Ausfallquote bei bis zu 71% (Statistisches Bundesamt, 2012).

Die tagebuchgestützten Erhebungen der amtlichen Statistik in Deutschland, beispielsweise die Einkommens- und Verbrauchsstichprobe (EVS), sind Quotenstichproben, die auf freiwilliger Basis durchgeführt werden.¹ Teilnahmebereite Haushalte werden über unterschiedliche direkte und indirekte Werbemaßnahmen gesucht und können sich für die Teilnahme melden. Aus allen

angemeldeten Haushalten wird nach einem vorgegebenen Quotenplan die Stichprobe gezogen. Da somit bei einer Quotenstichprobe der Kreis der potenziell für die Stichprobe vorgesehenen Haushalte offen ist, kann keine Responserate ermittelt werden. Stattdessen ist hier die Höhe der Quotenerfüllung ein Indikator für die Teilnahmebereitschaft. Seit Jahren zeigt sich eine rückläufige Entwicklung, wie das Beispiel der EVS als größte freiwillige Haushaltserhebung der amtlichen Statistik in Deutschland verdeutlicht: Wurde die Ist-Soll-Quote bei der zweiten EVS im Jahr 1969 noch zu rund 90% erfüllt, ist sie seit den 1980er-Jahren mit Ausnahme der EVS 1998 sukzessive gesunken und lag bei der EVS 2013 bei 67%. [↘ Tabelle 1](#)

Tabelle 1
Stichprobengrößen und Quotenerfüllung der Einkommens- und Verbrauchsstichproben

	Stichprobensoll	Nettostichprobe ¹	Ist-Soll-Quote
	Anzahl		%
Durchführung im früheren Bundesgebiet			
1962/63	52 857	38 110	72
1969	50 602	47 368	94
1973	52 238	48 841	93
1978	54 922	47 899	87
1983	54 944	44 507	81
1988	57 395	45 085	79
Durchführung im vereinten Deutschland			
1993	69 699	49 959	72
1998	73 890	62 150	84
2003	74 600	53 432	72
2008	77 648	55 110	71
2013	79 287	53 490	67

¹ Umfasst die Anzahl der Haushalte mit Jahresanschriftung/Haushaltsbuch.

Die sinkende Quotenerfüllung resultiert zum Teil aus einer steigenden Abbruchrate. Als Abbrüche gewertet werden Haushalte, die sich zunächst bereit erklären, an der EVS teilzunehmen, dann jedoch entweder gar nicht an der Erhebung teilnehmen oder zwar einzelne Erhebungsunterlagen bearbeiten, jedoch nicht das Haushaltsbuch führen.² Während die Abbruchrate bei der EVS 2008 bei 24% lag, stieg sie bei der EVS 2013 auf 27%. Für die EVS 2018 steht die Abbruchrate der-

¹ Weitere tagebuchgestützte, als Quotenstichproben umgesetzte Erhebungen sind die Laufenden Wirtschaftsrechnungen (Unterstichprobe der EVS) und die Zeitverwendungserhebung.

² Sie gelten als Abbrüche für die Haushaltsbuch-Stichprobe. Sofern die Erhebungsunterlagen „Allgemeine Angaben“ und gegebenenfalls „Geld- und Sachvermögen“ bearbeitet wurden, werden diese Angaben in den jeweiligen separaten Auswertungen berücksichtigt.

zeit noch nicht fest. Es zeigte sich jedoch bereits während der Erhebungsphase eine vergleichsweise negative Entwicklung, sodass mit einer höheren Abbruchrate als 2013 gerechnet wird.

Das Phänomen der sinkenden Teilnahmebereitschaft betrifft demnach nicht nur Deutschland oder spezielle Erhebungen. Eine zunehmende Verweigerungsrate führt ebenso zu geringeren Ausschöpfungsquoten wie eine steigende Abbruchrate. Folglich sind potenzielle Einflussfaktoren auf die Teilnahmebereitschaft zu identifizieren, um geeignete Maßnahmen einzuleiten. Deshalb werden nachfolgend zunächst Einflussfaktoren sowie dazugehörige Erkenntnisse aus der amtlichen Statistik vorgestellt. Im Anschluss folgt eine Beschreibung unterschiedlicher Maßnahmen zur Steigerung der Teilnahmebereitschaft an Haushaltserhebungen der amtlichen Statistik.

2

Einflussfaktoren auf die Teilnahmebereitschaft

Ob sich eine Person für oder gegen eine Teilnahme an einer Erhebung entscheidet und ob sie eine Erhebung vollständig durchführt oder vorzeitig abbricht, wird jeweils als Resultat einer Entscheidungsfindung beschrieben (Groves und andere, 1992). In der Regel wird davon ausgegangen, dass bei der Entscheidungsfindung Kosten und Nutzen beziehungsweise Vor- und Nachteile antizipiert und bewertet werden und daraus eine Handlung – Verweigerung, Teilnahme, Abbruch – resultiert. Der Entscheidungsprozess erfolgt dabei auf Basis von Faustregeln, sogenannten Heuristiken. Dabei wird eine oberflächliche und schnelle Kosten-Nutzen-Analyse durchgeführt, die von vielen externen, situativen und psychologischen (Interaktions-)Faktoren beeinflusst wird:

- › Faktoren auf der gesellschaftlichen Ebene, zum Beispiel die Legitimität der durchführenden Institution, die Akzeptanz von Erhebungen und deren Ergebnissen oder der Gesamtumfang an durchgeführten Befragungen (sogenannter Oversurveying-Effekt)
- › Merkmale des Befragungsdesigns, wie Erhebungsmodus, Befragungslänge oder -thema

- › Merkmale der Befragungsperson, beispielsweise soziodemografische Merkmale, gesundheitlicher Status, Wohnumfeld oder Vorerfahrungen mit Befragungen
- › Merkmale der Interviewerin oder des Interviewers (sofern vorhanden), zum Beispiel soziodemografische Merkmale, Erfahrungen, Erwartungen, Motivation oder emotionaler Status
- › Interaktion zwischen Interviewerin beziehungsweise Interviewer (sofern vorhanden) und Befragungsperson, die durch das jeweilige Verhalten, die jeweiligen Einstellungen und Erwartungen sowie durch situative Faktoren geprägt wird

Insbesondere zu den ersten drei Faktoren werden im Folgenden anhand einzelner Beispiele vorliegende Erkenntnisse im Zusammenhang mit den Haushaltserhebungen der amtlichen Statistik dargestellt.

Ergebnisse der Reputations- und Akzeptanzanalyse, die 2013 für das Statistische Bundesamt durchgeführt wurde (Lauff/Wachenfeld, 2014), geben Hinweise zu den Faktoren auf gesellschaftlicher Ebene. Demnach kannten 82 % der dort befragten Privatpersonen das Statistische Bundesamt zumindest vom Namen her. Rund 60 % schätzten es als vertrauenswürdig ein (Top-3-Wert auf einer siebenstufigen Zustimmungsskala³). Ein weiteres Ergebnis war, dass den Mikrozensus oder den Zensus lediglich jede vierte befragte Privatperson zumindest vom Namen her kannte.⁴ Die spontanen Assoziationen derjenigen, denen der Mikrozensus oder der Zensus bekannt war, umfassten in erster Linie eher allgemeine Begriffe wie „Volkszählung“ oder „Ermittlung der Bevölkerungszahl“. Fast jeder zehnten Person fiel spontan zu diesen Erhebungen gar nichts ein. Das Vertrauen gegenüber den Statistischen Ämtern des Bundes und der Länder und die Bekanntheit der Institutionen sowie der konkreten Erhebungen wirken sich auf die Akzeptanz der Datenerhebung aus. Aus der Reputations- und Akzeptanzanalyse ging ebenfalls hervor, dass zwar die Mehrheit der befragten Privatpersonen

3 Die Zustimmungsskala umfasste Antworten von 1 „trifft überhaupt nicht zu“ bis 7 „trifft voll und ganz zu“. Der Top-3-Wert umfasst somit die Summe der drei positiven Antwortoptionen 5 bis 7.

4 Die Bekanntheit der anderen Haushaltserhebungen der amtlichen Statistik, zum Beispiel die Einkommens- und Verbrauchsstichprobe, wurde nicht abgefragt. Es ist jedoch davon auszugehen, dass diese aufgrund ihres deutlich geringeren Umfangs und der Periodizität einen niedrigeren Bekanntheitswert erreichen.

eine Datenerhebung durch das Statistische Bundesamt als öffentliche Einrichtung akzeptierte oder dieser neutral gegenüberstand. Jedoch zeigte ein Teil der Befragten eine geringe oder keine Akzeptanz für die Datenerhebung. Bei den Erhebungen Mikrozensus und Zensus stellt die gesetzlich angeordnete Auskunftspflicht einen weiteren Faktor auf der gesellschaftlichen Ebene dar, der den Entscheidungsprozess für oder gegen eine Teilnahme beeinflusst.

Inwieweit Merkmale des Befragungsdesigns die Teilnahmeentscheidung beeinflussen können, geht beispielsweise aus einer von den Statistischen Ämtern des Bundes und der Länder im Jahr 2017 durchgeführten Testerhebung hervor. Mit ihr sollten Erkenntnisse zur Weiterentwicklung der Wirtschaftsrechnungen – bestehend aus der EVS und den Laufenden Wirtschaftsrechnungen – gewonnen werden. Dabei wurden von den zufällig ausgewählten Haushalten, die die freiwillige Teilnahme an der zwei- beziehungsweise vierwöchigen (je nach Testbedingung) tagebuchgestützten Erhebung ablehnten, unter anderem ihre Gründe für die Nichtteilnahme erfragt. Ebenso wurden Haushalte, die sich zunächst zur Teilnahme an der Testerhebung bereit erklärten, diese aber abgebrochen haben, nach ihren Abbruchgründen gefragt. In beiden Fällen wurde der zu hohe (Zeit-)Aufwand am häufigsten als Begründung genannt. In der Testerhebung fiel zudem die Abbruchrate bei einer zweiwöchigen Haushaltsbuchführung deutlich geringer aus als bei einer vierwöchigen Haushaltsbuchführung.

Testerhebung zur Weiterentwicklung der Wirtschaftsrechnungen 2017

Die Testerhebung nach §7 Absatz 2 Bundesstatistikgesetz diente unter anderem dazu, für die Wirtschaftsrechnungen Responseraten bei einer Zufallsstichprobe sowie für unterschiedliche Anschreibeziträume zu ermitteln. Auch Auswirkungen neu konzipierter Erhebungunterlagen sowie einer neuen Erhebungsmethodik (insbesondere die Verkürzung der Anschreibedauer und die Einführung retrospektiver Fragen) auf die Datenqualität waren zu untersuchen. Die Haushalte wurden für die freiwillige Testerhebung im Anschluss an die Mikrozensus-Befragung von den Interviewerinnen und Interviewern rekrutiert. Teilnehmende Haushalte führten unter anderem für einen Zeitraum von zwei beziehungsweise vier Wochen (je nach Testbedingung) ein Haushaltsbuch. Für die vollständige Teilnahme erhielten die Haushalte eine Geldprämie in Höhe von 70 Euro.

Weiterhin geht aus den bei der Testerhebung ebenfalls erfassten Teilnahmegründen hervor, dass auch Merkmale der Befragungspersonen einen Einfluss auf die Teilnahmeentscheidung haben. So gaben Haushalte mit einem höheren Einkommen an, vor allem teilgenommen zu haben, weil es sich um eine amtliche Statistik handelt. Die Geldprämie wurde insbesondere von Haushalten im unteren Einkommensbereich und von Haushalten mit Kindern als wichtiger Teilnahmegrund genannt. Hinzu kommt, dass auch die Abbruchrate bei der EVS nicht bei allen Haushalten gleich ausfällt. Haushalte mit Kindern, mit geringem Einkommen, von Nichterwerbstätigen (ausgenommen Personen im Ruhestand), Landwirtinnen und Landwirten sowie von Selbstständigen wiesen bei der EVS 2013 mit rund 30 bis 40 % überdurchschnittlich hohe Abbruchraten auf.

Eine systematische Analyse der Auswirkungen unterschiedlicher Merkmale von Interviewerinnen und Interviewern auf die Teilnahmebereitschaft von amtlichen Haushaltserhebungen liegt nicht vor. Dass jedoch eine persönliche Ansprache einen wichtigen Einfluss haben kann, ist ebenfalls aus der Testerhebung abzuleiten. Dort wurden Haushalte im Anschluss an die Mikrozensus-Befragung von den Interviewerinnen und Interviewern rekrutiert. Die Entscheidung zur Teilnahme wurde ähnlich häufig damit begründet, dass

- › das persönliche Gespräch mit der Interviewerin beziehungsweise dem Interviewer überzeugt hat,
- › es sich um eine amtliche Statistik handelt oder
- › es für die Teilnahme eine Geldprämie gab.

Diese Erkenntnisse verdeutlichen, dass sich vielfältige Faktoren auf den Entscheidungsprozess für oder gegen eine Teilnahme an einer Erhebung auswirken können. Somit sind auch vielfältige Maßnahmen möglich, um die Entscheidung für eine Teilnahme positiv zu beeinflussen. Im Folgenden werden bezogen auf die freiwilligen Haushaltserhebungen der amtlichen Statistik entsprechende Maßnahmen vorgestellt.

3

Maßnahmen zur Steigerung der Teilnahmebereitschaft

Das als Tailored-Design-Methode (Dillman und andere, 2009) bekannte Konzept beschreibt drei globale Maßnahmen, um die Teilnahmebereitschaft an Erhebungen zu steigern:

- › Steigerung des Teilnahmenutzens
- › Senkung der Teilnahmekosten
- › Aufbau von Vertrauen

Nach der Tailored-Design-Methode sowie den Erfahrungen der amtlichen Statistik sind nicht nur einzelne Maßnahmen erforderlich, um die Teilnahmebereitschaft zu steigern und vorzeitige Abbrüche zu reduzieren. Vielmehr bedarf es eines Bündels aus unterschiedlichen Maßnahmen. Dabei sind sowohl erhebungsspezifische als auch erhebungsübergreifende Maßnahmen einzubeziehen und zielgruppenspezifische Auswirkungen sowie die jeweiligen Rahmenbedingungen der einzelnen Haushaltserhebungen zu berücksichtigen.

3.1 Steigerung des Teilnahmenutzens

Studien weisen darauf hin, dass die Ankündigung, am Ende der Teilnahme basierend auf den eigenen Angaben bei der Erhebung ein persönliches Ergebnis zu erhalten, die Teilnahmebereitschaft erhöht (Groves und andere, 1992). Eine individuelle Ergebnisübersicht bietet beispielsweise das elektronische Haushaltsbuch in den Laufenden Wirtschaftsrechnungen. Es enthält eine Auswertungsfunktion, die dem Haushalt (grafische) Übersichten der eigenen Einnahmen und Ausgaben nach Kategorien anzeigt. Die Möglichkeit, die individuellen Ergebnisse mit den Ergebnissen anderer Haushalte vergleichen zu können, kann einen zusätzlichen Teilnahmenutzen schaffen. Dabei ist insbesondere ein Vergleich mit ähnlichen Personen oder Gruppen nützlich, um die eigenen Ergebnisse realistisch bewerten und einordnen zu können. Rückmeldungen von Teilnehmenden der EVS 2013 zeigen, dass ein Vergleich der eigenen Ausgaben für Nahrungsmittel, Bekleidung und vieles mehr mit dem Gesamtdurchschnitt oder dem Durchschnittswert ähnlicher Haushalte gewünscht ist. Für die EVS 2023 wird

ein digitales Erhebungsinstrument entwickelt. Dabei wird erwogen, individuelle Auswertungs- beziehungsweise Übersichtsfunktionen sowie Funktionen, die einen Vergleich der eigenen Ausgaben mit denen von anderen Haushalten ermöglichen, zu integrieren. Damit wird zugleich die Attraktivität für die Nutzung des digitalen Erhebungsinstruments anstelle eines Papierfragebogens erhöht.

Ein Teilnahmenutzen besteht auch darin, mit den Angaben einen gesellschaftlichen Beitrag beziehungsweise einen Beitrag zu etwas Wichtigem zu leisten. Deshalb wird zum Beispiel im Werbematerial für die Teilnahme an der EVS betont, dass die Teilnahme einen wichtigen Beitrag zur realistischen Beurteilung der Lebensbedingungen in Deutschland leistet. Das Herausstellen dieser eher altruistischen Motivatoren in Werbemitteln oder Anschreiben ist auch künftig wichtig. Hierdurch können Haushalte für die Teilnahme gewonnen werden, für die andere Anreize (Incentives) eine eher geringe Bedeutung haben (zum Beispiel einkommensstarke Haushalte).

Aktuell erhalten Haushalte, die vollständig an einer freiwilligen Erhebung wie der EVS teilgenommen haben, eine Geldprämie (monetäres Postpaid-Incentive). Die Studienlage (zum Beispiel Göritz, 2014; Engel/Schnabel, 2004) sowie die Erfahrungen der Statistischen Ämter des Bundes und der Länder mit freiwilligen Haushaltserhebungen der amtlichen Statistik zeigen eindeutig, dass monetäre Incentives die Teilnahmebereitschaft erhöhen. Deshalb ist bei freiwilligen Haushaltserhebungen die Auszahlung von Aufwandsentschädigungen weiterhin unverzichtbar. Auch Verlosungen, insbesondere von Geldbeträgen, können einen positiven Beitrag auf die Teilnahmebereitschaft haben. So setzt das Statistikamt der Niederlande Verlosungen aufgrund ihrer positiven Wirkung auf die Teilnahmebereitschaft als Standard-Incentive ein (Luiten, 2016). Deshalb ist zu überlegen, künftig zusätzlich zur sicheren Teilnahmeprämie auch Verlosungen durchzuführen.

Ein konkreter Geldbetrag war aufgrund der je Bundesland unterschiedlichen Prämienhöhen in bundesweiten Werbemaßnahmen zur EVS 2018 nicht zu benennen. Angaben über die Höhe der Geldbeträge wurden nur in länderspezifischen Werbemitteln ausgewiesen, teilweise jedoch eher unauffällig. Hierdurch wurde der hohe Nutzen der Geldprämie als Teilnahmeanreiz – insbesondere für schwer anzuwerbende Gruppen, beispielsweise

einkommensschwache Haushalte – nicht voll ausgeschöpft. Die unterschiedlich hohen Geldprämien führten zudem beispielsweise auf Online-Plattformen zu Diskussionen zwischen potenziellen Auskunftgebenden. Um dem entgegenzuwirken, soll bei der EVS 2023 in allen Bundesländern eine identische Teilnahmeprämie ausbezahlt werden.

Beim Einsatz der unterschiedlichen Incentives ist zu berücksichtigen, dass deren Einfluss auf die Teilnahmeentscheidung auch davon abhängen kann, wie sie kommuniziert und folglich seitens der Teilnehmenden wahrgenommen werden. So löst die Beschreibung „Als Aufwandsentschädigung erhalten Sie 100 Euro“ gedanklich eine ökonomische Kosten-Nutzen-Analyse aus. Dabei wird der voraussichtliche Zeitaufwand (zum Beispiel die dreimonatige Haushaltsbuchführung) ins Verhältnis zur Entschädigung (100 Euro) gesetzt, die dann gegebenenfalls als zu niedrig bewertet wird. Die Beschreibung „Als Dankeschön für Ihre Teilnahme erhalten Sie 100 Euro“ ändert zwar nicht die tatsächliche Höhe der Geldprämie, kann jedoch die Wahrnehmung und kognitive Beurteilung verändern. Ähnliche positive Effekte können auch durch andere Formulierungen erzielt werden. So ist eine Anfrage zur Teilnahme an einer Befragung erfolgreicher, wenn diese Anfrage als eine Bitte um Mithilfe beschrieben wird (Dillman und andere, 2014). Wird die Teilnahme an einer Befragung als eine „seltene Chance“ oder „etwas Besonderes“ kommuniziert, weil der Teilnehmerkreis begrenzt ist oder der Haushalt zu den zufällig „Ausgewählten“ gehört, erhöht sich die Teilnahmebereitschaft (Groves und andere, 1992).

3.2 Senkung der Teilnahmekosten

Verschiedene Aspekte, wie die aufzuwendende Zeit, die Anzahl, Komplexität und Inhalte der Fragen, die Erhebungswege oder die Art und der Umfang an Unterstützungsmöglichkeiten, bestimmen die Teilnahmekosten. Insbesondere die EVS und die Laufenden Wirtschaftsrechnungen sind aufwendige Haushaltserhebungen, bei denen Teilnehmende unter anderem für einen vorgegebenen Zeitraum sämtliche Einnahmen und Ausgaben in ein Haushaltsbuch eintragen. Bis einschließlich 1993 erfolgte bei der EVS die Haushaltsbuchführung für einen Zeitraum von zwölf Monaten (Jahresanschreibung). Ab der EVS 1998 wurde

der Zeitraum auf drei Monate verkürzt (quartalsweise Anschreibung).¹⁵ Diese methodische Umstellung ging zunächst mit einer höheren Quotenerfüllung einher. Das Beispiel zeigt aber auch, dass der positive Effekt nachlassen kann: Bei der EVS 2003 lag die Ist-Soll-Quote wieder auf dem gleichen Niveau wie bei der EVS 1993. Dies deutet darauf hin, dass verschiedene Maßnahmen erforderlich sind und Anstrengungen in anderen Bereichen nicht reduziert werden sollten.

Eine umfassende Betreuung beziehungsweise Unterstützung der Auskunftgebenden kann insbesondere bei den tagebuchgestützten Erhebungen die wahrgenommenen Teilnahmekosten und somit die Abbruchwahrscheinlichkeit reduzieren. Informationen über den Erhebungsablauf, das Ausfüllen der Erhebungsunterlagen oder den Datenschutz werden den Teilnehmenden postalisch bereitgestellt. Ebenso stehen ihnen bei Fragen Ansprechpersonen telefonisch oder per E-Mail zur Verfügung. Ergänzend zu diesen passiven Betreuungsangeboten werden bei den tagebuchgestützten Erhebungen vereinzelt auch aktive Betreuungsmaßnahmen umgesetzt, um Haushalte für die (weitere) Teilnahme zu motivieren und Abbrüche zu vermeiden. Hierzu zählen telefonische oder E-Mail-Erinnerungen kurz vor Beginn sowie kurz vor Ende des Teilnahmezeitraums oder auch Betreuungs- beziehungsweise Motivationstelefonate während des Teilnahmezeitraums. Insbesondere die telefonischen Kontakte sind wichtig. Durch das persönliche Gespräch erfahren die Teilnehmenden Wertschätzung, gegebenenfalls bestehende Hemmschwellen – sich etwa bei Fragen an die Hotline zu wenden – werden reduziert. Deshalb ist für die EVS 2023 eine Ausweitung der Betreuung vorgesehen, insbesondere auch in Form aktiver telefonischer Kontakte.

Werden das Angebot an Erhebungswegen beziehungsweise Teilnahmemöglichkeiten als zu aufwendig, nicht intuitiv, zu eingeschränkt oder als Eingriff in die Privatsphäre empfunden, kann dies die Teilnahmekosten erhöhen. Ein Angebot an unterschiedlichen Erhebungswegen (persönlich, telefonisch, schriftlich, online),

5 Die Verkürzung der Anschreibedauer zur EVS 1998 war auch mit einer veränderten Anschreibung der Ausgaben verbunden. Vor 1998 wurden ausgewählte Ausgaben, zum Beispiel für Wohnen, Energie, Bekleidung und Möbel, während des gesamten Jahres erhoben. Ausgaben für Nahrungsmittel, Getränke und Tabakwaren sowie weitere regelmäßig getätigte Ausgaben wurden hingegen nur für einen Monat erfasst. Ab 1998 wurden alle Ausgaben für einen Zeitraum von drei Monaten erfasst.

aus denen die Teilnehmenden ihren Fähigkeiten und Präferenzen entsprechend wählen können, senkt diese Teilnahmekosten. Die freiwilligen Haushalts-erhebungen werden derzeit überwiegend schriftlich erhoben. Digitale Erhebungsinstrumente werden vereinzelt eingesetzt – für den Erhebungsteil Allgemeine Angaben der EVS sowie das Haushaltsbuch bei den Laufenden Wirtschaftsrechnungen. Online-Befragungen beziehungsweise digitale Erhebungsinstrumente werden seitens teilnehmender Haushalte zunehmend erwartet und sind insbesondere als Alternative zum Papierfragebogen unverzichtbar. Deshalb ist für die EVS 2023 die Einführung eines digitalen Erhebungsinstruments vorgesehen. Der digitale Erhebungsweg soll das vorhandene Angebot ergänzen und Teilnahmebarrieren für bestimmte Zielgruppen, beispielsweise jüngere oder technikaffine Personen, senken.

3.3 Aufbau von Vertrauen

Widerstand gegenüber verpflichtenden Haushalts-erhebungen oder eine geringe Bereitschaft, an freiwilligen Erhebungen teilzunehmen, können verschiedene Gründe haben. Beispiele dafür sind ein unzureichendes Vertrauen in die Einhaltung des Datenschutzes oder eine geringe Akzeptanz der Auskunftspflicht. Um dem entgegenzuwirken, sind entsprechende Kommunikationsmaßnahmen unerlässlich. Für den Zensus 2011 wurden zahlreiche Informationsmaßnahmen durchgeführt, beispielsweise Berichte und Interviews für die (über-)regionale Presse, Erläuterungsvideos oder eine Homepage. Diese breite Öffentlichkeitsarbeit wirkte sich sogar auf den folgenden Mikrozensus positiv aus. Ähnliches war auch in den Vereinigten Staaten beim Zensus 2010 zu beobachten, als das U.S. Census Bureau eine bis dahin einmalig umfangreiche Kommunikationskampagne durchgeführt hat. Damit konnten Bekanntheit, Einstellung und Wissen über den Zensus verbessert werden (Datta und andere, 2012). Bei der EVS 2018 wurden zusätzlich zu Plakaten und Flyern erstmals digitale Werbemittel [Videoclip auf YouTube (www.youtube.com/user/destatis), Internetbanner] erstellt, um Haushalte auch über soziale Medien und andere Internetkanäle ansprechen zu können. Die ausweiteten Werbemaßnahmen führten dazu, dass sich im Vergleich zur EVS 2013 mehr Haushalte zur Teilnahme bereit erklärten. Basierend auf diesen ersten Erfahrun-

gen sollte überlegt werden, künftig solche Maßnahmen stärker einzusetzen, um damit die Teilnehmerzahl derjenigen Haushalte zu erhöhen, die in der Gesamtstichprobe eher schwach vertreten sind. Die angeführten Beispiele zeigen, dass zielgruppenspezifische Kommunikationsmaßnahmen auf unterschiedlichen Wegen die Akzeptanz beziehungsweise Teilnahmebereitschaft positiv beeinflussen. Da sich der Erfolg solcher Maßnahmen über die Zeit verliert, müssen sie kontinuierlich durchgeführt werden, um langfristig und nachhaltig die Teilnahmebereitschaft in den Haushaltserhebungen zu sichern.

Eine Befragung von Privatpersonen im Rahmen der Reputations- und Akzeptanzanalyse für das Statistische Bundesamt im Jahr 2013 ergab, dass die Bereitstellung von Ergebnissen von rund jeder fünften Befragungsperson als Basisfaktor und von fast jeder vierten Person als Begeisterungsfaktor⁶ gewertet wird (Lauff und andere, 2014). Dabei bedeutet Basisfaktor, dass eine solche Leistung erwartet wird und seine Nichterfüllung zu Verärgerung führt. Begeisterungsfaktor bedeutet, dass diese Leistung nicht explizit erwartet wird und die Erfüllung Begeisterung hervorruft. Obwohl alle Ergebnisse veröffentlicht werden und der breiten Öffentlichkeit zur Verfügung stehen, werden den Auskunftgebenden nur vereinzelt ausgewählte Ergebnisse direkt zur Verfügung gestellt. Einige Statistische Landesämter bedanken sich postalisch bei den Teilnehmerhaushalten mit ausgewählten Ergebnissen der EVS. Um das Vertrauen in die amtliche Statistik zu stärken und die Akzeptanz der Datenerhebung zu erhöhen, ist es empfehlenswert, die direkte Ergebnisbereitstellung auszuweiten, optional auch über E-Mail.

Zu den direkten Werbemaßnahmen, um teilnahmebereite Haushalte für die freiwilligen Haushaltserhebungen zu gewinnen, zählen unter anderem postalische Anschreiben. Studien zeigen, dass Anschreiben einen hohen Stellenwert beim Aufbau von Vertrauen und bei der Entscheidung für oder gegen eine Teilnahme haben (siehe zum Beispiel Porst, 2001; Edwards und andere, 2002, 2009). Es wird beispielsweise empfohlen, dass der Briefumschlag einen seriösen Eindruck vermittelt,

⁶ Basis- und Begeisterungsfaktoren entsprechend dem Kano-Modell. Dieses beschreibt den Zusammenhang zwischen dem Erfüllungsgrad bestimmter Produkteigenschaften und der Kundenzufriedenheit mit dem Produkt (Kano und andere, 1984).

die Anrede persönlich ist und das Anschreiben maximal eine Seite umfasst. Die Beigabe kleiner Geschenke (sogenannte Prepaid-Incentives) sowie Unterstützungsschreiben von anerkannten Autoritäten¹⁷ schaffen zusätzlich Vertrauen und erhöhen die Teilnahmebereitschaft. Welchen Einfluss Veränderungen am Anschreiben haben können, zeigen die Erfahrungen von Statistik Austria bei der Konsumerhebung 2014/15¹⁸ (Riegler, 2016). Aufgrund einer zunächst niedrigen Online-Registrierungsrate wurde der Anschreibetext gekürzt und zielte in der neuen Version stärker darauf ab, den Zweck der Erhebung und die Wichtigkeit der Teilnahme möglichst greifbar und anschaulich darzustellen. Vor den Anpassungen registrierten sich 3% der Angesprochenen für die Teilnahme online und 9% füllten alle drei Befragungsteile aus. Nach den Anpassungen betrug die Online-Registrierungsrate 9% und der Anteil der vollständigen Teilnahmen stieg auf 33%. Aufgrund dieser Erkenntnisse wurde für die EVS 2023 die Überarbeitung insbesondere der postalischen Werbeanschreiben angestoßen.

Zahlreiche Studien zeigen, dass die bereits erwähnten Prepaid-Incentives (bedingungslos zu Beginn überreichte, meist kleine Geschenke) neben Belohnungen, die Teilnehmende nach Erhebungsabschluss erhalten (Postpaid-Incentives), als vertrauensbildende Maßnahme einen positiven Einfluss auf die Teilnahmebereitschaft haben (siehe zum Beispiel Becker und andere, 2007; Church, 1993; Dillman und andere, 2009; Göritz, 2015; Lund/Golloy, 2016; Luiten, 2016). Hierbei erzeugt eher der Zeitpunkt der Vergabe (vor der Teilnahme) als der monetäre Wert des Incentives eine positive Wirkung. Diese fällt insbesondere in Kombination mit einem Postpaid-Incentive höher aus als die alleinige Vergabe eines Postpaid-Incentives. Die bedingungslos überreichten Geschenke eröffnen eine soziale Austauschbeziehung und fungieren als ein „Symbol des Vertrauens“. Bei den Beschenkten bewirkt das Geschenk ein Gefühl der sozialen Verpflichtung, der Bitte nach einer Teilnahme an der Erhebung zu entsprechen. Bislang lie-

gen keine Erkenntnisse über die konkrete Wirkung von Prepaid-Incentives auf die Teilnahmebereitschaft bei Haushaltserhebungen der amtlichen Statistik vor. Die positiven Ergebnisse zahlreicher anderer Anwendungsfälle sollten jedoch zum Anlass genommen werden, künftig die Wirkung kleiner Beigaben auf die Anmeldequote zumindest zu testen – beispielsweise bei der postalischen Anwerbung für die EVS.

4

Fazit

Die sinkende Quotenerfüllung bei tagebuchgestützten Haushaltserhebungen wie der EVS sowie die zum Teil geringen Responderaten beim Mikrozensus für Merkmale ohne Auskunftspflicht beeinträchtigen die Datenqualität dieser Erhebungen. Sie erfordern den Einsatz geeigneter Maßnahmen, um die Teilnahmebereitschaft zu steigern und Verweigerungen sowie Abbrüche zu reduzieren. Dabei ist nicht nur die Umsetzung einzelner Maßnahmen erforderlich, sondern vielmehr ein Bündel aus unterschiedlichen Maßnahmen. Diese zielen darauf ab, den Teilnahmenutzen zu erhöhen, die Teilnahmekosten zu senken und das Vertrauen zu stärken:

- › Den Teilnahmenutzen erhöhen Anreize wie das Bereitstellen individueller Ergebnisaufbereitungen, monetäre Incentives oder der Hinweis, wie mit der eigenen Teilnahme ein Beitrag für die Gesellschaft geleistet werden kann.
- › Die Teilnahmekosten können gesenkt werden, indem die aufzuwendende Zeit möglichst gering ausfällt, Teilnehmende betreut und unterstützt werden und unterschiedliche Erhebungswege angeboten werden.
- › Das Vertrauen kann durch zielgruppenspezifische Kommunikationsmaßnahmen, die Berücksichtigung der Empfehlungen für postalische Anschreiben, bedingungslose kleine Anerkennungen, die Bereitstellung von Ergebnissen sowie anschauliche Erklärungen zur Methodik gestärkt werden.

Bei der EVS 2018 wurden mit der Nutzung digitaler Werbemittel bereits Maßnahmen zur Steigerung der Teilnahmebereitschaft eingesetzt. Daran anknüpfend

7 So wurde bei der Studie „Alterssicherung in Deutschland“ (ASID) eine verbesserte Responderate unter anderem auf das eingesetzte Versandkonzept mit einem von Staatssekretär Albrecht unterzeichneten Ankündigungsschreiben zurückgeführt (Heien/Heckmann, 2017). Ähnliches gilt für Unterstützungsschreiben von Bürgermeisterinnen und Bürgermeistern im Rahmen der Nationalen Verzehrsstudie (NVZ) II (Max-Rubner-Institut, 2008).

8 Die Konsumerhebung in Österreich ist vergleichbar mit der deutschen EVS.

wird zur EVS 2023 ein digitales Erhebungsinstrument eingeführt und eine bundesweit einheitliche Teilnahmeprämie umgesetzt. Für Haushaltserhebungen mit Auskunftspflicht wie den Mikrozensus ist zur Reduzierung von Erhebungswiderständen zu prüfen, inwieweit auf ähnliche Maßnahmen zurückgegriffen werden kann und inwieweit die Entwicklung neuer Maßnahmen erforderlich ist. [\[1\]](#)

LITERATURVERZEICHNIS

Becker, Rolf/Imhof, Regula/Mehlkop, Guido. *Die Wirkung monetärer Anreize auf den Rücklauf bei einer postalischen Befragung und die Antworten auf Fragen zur Delinquenz – Empirische Befunde eines Methodenexperiments*. In: Methoden, Daten, Analysen (mda). Jahrgang 1. Ausgabe 2/2007, Seite 131 ff.

Church, Allan H. *Estimating the effect of incentives on mail survey response rates: a meta-analysis*. In: Public Opinion Quarterly. Jahrgang 57. Ausgabe 1/1993, Seite 62 ff.

Datta, Rupa A./Yan, Ting/Evans, Doug/Pedlow, Steven/Spencer, Bruce/Bautista, Rene. *2010 Census Integrated Communications Program Evaluation (CICPE) – Final Report*. In: 2010 Census Planning Memoranda Series. No. 167. 2012.

De Leeuw, Edith/de Heer, Wim. *Trends in Household Survey Nonresponse: A Longitudinal and International Comparison*. In: Groves, Robert M./Dillman, Don A./Eltinge, John L./Little, Roderick J. A. (Herausgeber). *Survey Nonresponse*. New York 2002, Seite 41 ff.

Dillman, Don A./Smyth, Jolene D./Christian, Leah Melani. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 3. Auflage. New York 2008.

Dillman, Don A./Smyth, Jolene D./Christian, Leah Melani. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 4. Auflage. New York 2014.

Edwards, Philip James/Roberts, Ian/Clarke, Mike J./DiGuseppi, Carolyn/Pratap, Sarah/Wentz, Reinhard/Kwan, Irene. *Increasing response rates to postal questionnaires: systematic review*. In: BMJ. Band 324. Nummer 7347/2002, Seite 1183 ff.

Edwards, Philip James/Roberts, Ian/Clarke, Mike J./DiGuseppi, Carolyn/Wentz, Reinhard/Kwan, Irene/Cooper, Rachel/Felix, Lambert M./Pratap, Sarah. *Methods to increase response to postal and electronic questionnaires*. In: Cochrane Database Syst Revue. Ausgabe 3/2009, Seite 1 ff.

Engel, Uwe/Schnabel, Christiane. *Markt- und Sozialforschung – Metaanalyse zum Ausschöpfungsgrad*. ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute. 2004.

Göritz, Anja S. *Incentive effects*. In: Engel Uwe/Jann, Ben/Lynn, Peter/Scherpenzeel, Annette/Sturgis Patrick (Herausgeber). *Improving Survey Methods: Lessons from Recent Research*. New York 2015, Seite 339 ff.

Groves, Robert M./Cialdini, Robert B./Couper, Mick P. *Understanding the decision to participate in a survey*. In: Public Opinion Quarterly. Jahrgang 56. Ausgabe 4/1992, Seite 475 ff.

Heien, Thorsten/Heckmann, Jochen. *Alterssicherung in Deutschland 2015 (ASID 2015) – Methodenbericht*. Forschungsbericht 474/M des Bundesministeriums für Arbeit und Soziales. München 2017.

LITERATURVERZEICHNIS

Kano, Noriaki/Seraku, Nobuhiko/Takahashi, Fumio/Tsuji, Shin-ichi. *Attractive Quality and Must-Be Quality*. In: Journal of The Japanese Society for Quality Control. Jahrgang 14. Ausgabe 2/1984, Seite 147 ff.

Lauff, Helen/Wachenfeld, Alexandra. *Abschlussbericht zum Projekt Reputations- und Akzeptanzanalyse für das Statistische Bundesamt*. Frankfurt 2014.

Lund, Kari-Anne/Golloy, Elisabeth. *The use of incentives in household surveys: a future strategy to save response rates, or a relic of the past?* Beitrag auf der Nordiskt Statistekermöte. Stockholm 2016.

Luiten, Annemieke. *Incentives in Official Statistics: Effects on response, representativeness and target variables*. Beitrag auf dem 27. Internationalen Workshop on Household Survey Nonresponse. Oslo 2016.

Luiten, Annemieke/de Leeuw, Edith/Schouten, Barry. *First results of the International questionnaire on nonresponse*. Beitrag auf dem 27. Internationalem Workshop on Household Survey Nonresponse. Oslo 2016.

Max Rubner-Institut (Herausgeber). *Nationale Verzehrsstudie II: Die bundesweite Befragung zur Ernährung von Jugendlichen und Erwachsenen – Ergebnisbericht, Teil 1*. Karlsruhe 2008.

Porst, Rolf. *Wie man die Rücklaufquote bei postalischen Befragungen erhöht*. In: Zentrum für Umfragen, Methoden und Analysen. -ZUMA- GESIS, Leibniz-Institut für Sozialwissenschaften. Mannheim 2001.

Riegler, Romana. *The new CAWI mode and response rates of the Austrian HBS 2014/15*. Beitrag beim Treffen der Working Group „Statistics on Living Conditions / HBS“. Luxemburg 2016.

Simmons, Eleanor/Wilmot, Amanda. *Incentive Payments on Social Surveys: a Literature Review*. In: Survey Methodology Bulletin. Band 53/2004, Seite 1 ff.

Statistisches Bundesamt. *Qualitätsbericht Mikrozensus 2011*. Wiesbaden 2012.

HARMONISIERUNG DER UMSATZSTEUER-MELDEFRISTEN

Henrik Neumann, Kristina Trusheim

↘ **Schlüsselwörter:** Erfüllungsaufwand – Umsatzsteuer-Voranmeldung – Zusammenfassende Meldung – Umsatzsteuer-Meldefristen – Bürokratiekostenmessung

ZUSAMMENFASSUNG

Das Umsatzsteuergesetz regelt sehr ausführlich unter anderem das Besteuerungsverfahren und die damit verbundenen Fristen. Zu unterschiedlichen Zeitpunkten müssen umsatzsteuerpflichtige Unternehmen an verschiedene Finanzbehörden umfangreiche Erklärungen elektronisch übermitteln. Im europäischen Vergleich ist das deutsche Verfahrensrecht zur Umsatzsteuer besonders komplex gestaltet.

Das Bundeskanzleramt hat das Statistische Bundesamt beauftragt, die Auswirkungen einer Angleichung der Abgabezeitpunkte und Meldezeiträume der Umsatzsteuer-Voranmeldung und der Zusammenfassenden Meldung zu untersuchen. Der Aufsatz beschreibt das methodische Vorgehen bei dieser Untersuchung, spiegelt die wesentlichen Ergebnisse zum Status quo wider und zeigt die mit einer Harmonisierung der Umsatzsteuer-Meldefristen verbundenen Entlastungspotenziale auf.

↘ **Keywords:** *compliance costs – advance turnover tax return – recapitulative statement – turnover tax reporting deadlines – bureaucracy cost measurement*

ABSTRACT

The Turnover Tax Act comprehensively regulates, among other things, the taxation procedure and the associated reporting deadlines. Enterprises liable to pay turnover tax must electronically transmit comprehensive tax returns to various tax authorities at different times. Compared with other European countries, the German procedural regulations on turnover tax are particularly complex.

The Federal Chancellery has commissioned the Federal Statistical Office to examine the effects of aligning the submission times and reporting periods of the advance turnover tax returns and of the recapitulative statement. This article describes the study's methodological approach, presents the main status-quo results, and identifies the potential for burden reduction associated with a harmonisation of the turnover tax reporting deadlines.



Henrik Neumann

ist Politikwissenschaftler und Referent im Referat „Aufwandsermittlung, Kostenschätzung (Standort Wiesbaden)“ des Statistischen Bundesamtes. Er plant und koordiniert Projekte in ausgewählten Lebens- und Rechtsbereichen mit dem Ziel des Bürokratieabbaus, der besseren Rechtsetzung und der Identifizierung von Vereinfachungen.



Kristina Trusheim

ist Diplom-Kauffrau und Referentin im Referat „Aufwandsermittlung, Kostenschätzung (Standort Wiesbaden)“ des Statistischen Bundesamtes. Sie organisiert Nachmessungen des Erfüllungsaufwands zwei Jahre nach Inkrafttreten eines Gesetzes und betreut Projekte zur Ermittlung des bürokratischen Aufwands von Bürgern, Wirtschaft und Verwaltung.

1

Einleitung

1.1 Projekthintergrund

Im Februar 2018 führte das Bundeskanzleramt in Zusammenarbeit mit dem Statistischen Bundesamt in Berlin einen Expertenworkshop zur Vereinfachung des Umsatzsteuerverfahrens durch. Anlass dazu waren die Ergebnisse der Lebenslagenbefragung 2017 (Himmelsbach und andere, 2016), in der die Zufriedenheit der Bürgerinnen und Bürger sowie der Unternehmen mit Dienstleistungen der Verwaltung ermittelt worden war. Die befragten Unternehmen waren insbesondere mit der Verständlichkeit des (Steuer-)Rechts sowie der Verständlichkeit der Formulare und Anträge nur unterdurchschnittlich zufrieden. Um konkrete Probleme zu identifizieren und Lösungsvorschläge zu erarbeiten, beteiligten sich Vertreterinnen und Vertreter von Unternehmen, Verbänden, der Finanzverwaltung, des Bundesministeriums der Finanzen und des Bundesministeriums für Wirtschaft und Energie am Workshop. Von den Teilnehmerinnen und Teilnehmern wurde unter anderem kritisiert, dass die Abgabefristen der Umsatzsteuer-Voranmeldung und der Zusammenfassenden Meldung zeitlich auseinander driften, obwohl beide Meldungen im Wesentlichen auf derselben Datengrundlage basieren. Eine Angleichung der Abgabezeitpunkte verspräche nach Einschätzung der Teilnehmenden eine Erleichterung für die betroffenen Unternehmen.

1.2 Umsatzsteuer-Meldefristen

Die Umsatzsteuer-Jahreserklärung muss im Regelfall bis zum 31. Juli des Folgejahres elektronisch an das Betriebsstättenfinanzamt übermittelt werden [§ 18 Absatz 3 Umsatzsteuergesetz (UStG) in Verbindung mit § 149 Absatz 3 Abgabenordnung (AO)]. Daneben müssen Unternehmen mit einer verbleibenden Umsatzsteuer zwischen 1 000 und 7 500 Euro vierteljährlich eine Voranmeldung abgeben. Betrug die Umsatzsteuer im vorangegangenen Kalenderjahr mehr als 7 500 Euro, muss monatlich eine Umsatzsteuer-Voranmeldung erstellt werden (§ 18 Absatz 2 UStG). Die Abgabefrist endet jeweils am 10. Tag nach Ablauf des Meldezeitraums (§ 18

Absatz 1 UStG). Die Frist kann durch die Beantragung einer Dauerfristverlängerung um einen Monat verlängert werden (§ 46 Umsatzsteuer-Durchführungsverordnung – UStDV). Für Quartalsmelder gilt die Dauerfristverlängerung nach einmaliger Beantragung bis zum Widerruf fort. Monatsmelder müssen jeweils im Februar eines Jahres eine Sondervorauszahlung in Höhe von einem Elftel der Summe der Vorauszahlungen für das vorangegangene Kalenderjahr anmelden, um die Fristverlängerung nutzen zu können (§ 47 UStDV). Die Vorauszahlung wird mit der Umsatzsteuer-Voranmeldung für den Monat Dezember verrechnet.

Exportieren Unternehmen Waren oder Dienstleistungen in andere Mitgliedstaaten der Europäischen Union (EU), müssen sie darüber hinaus eine Zusammenfassende Meldung an das Bundeszentralamt für Steuern übermitteln (§ 18a UStG). Im Regelfall ist der Meldezeitraum der Zusammenfassenden Meldung der Kalendermonat. Sofern jedoch bestimmte Umsatzgrenzen nicht überschritten werden, kann die Zusammenfassende Meldung auch quartalsweise oder jährlich eingereicht werden. Bis 2010 musste die Zusammenfassende Meldung – wie auch die Umsatzsteuer-Voranmeldung – bis zum 10. Tag nach Ablauf des Meldezeitraums eingereicht werden. Nach der Umsetzung einer europarechtlichen Vorgabe weichen die Abgabetermine voneinander ab. Die Zusammenfassende Meldung muss in Deutschland nunmehr bis zum 25. Tag nach Ablauf des Meldezeitraums übermittelt werden. Die Richtlinie schreibt den Mitgliedstaaten außerdem vor, bei innergemeinschaftlichen Umsätzen den Finanzverwaltungen grundsätzlich monatlich Zusammenfassende Meldungen zur Verfügung zu stellen, um Steuerbetrug zu bekämpfen.

Im Vergleich mit anderen EU-Mitgliedstaaten ist das deutsche Verfahren besonders komplex gestaltet. In 13 anderen europäischen Ländern ist es gelungen, einen einheitlichen Abgabetermin für die Umsatzsteuer-Voranmeldung und die Zusammenfassende Meldung im letzten Drittel des auf den Meldezeitraum folgenden Monats zu etablieren. In Deutschland haben Unternehmen, die ihre Umsatzsteuer-Voranmeldung zum regulären Termin (10. Tag des Folgemonats) abgeben, innerhalb der EU die kürzeste Bearbeitungsfrist. Wird die Dauerfristverlängerung genutzt, verschiebt sich der Termin (10. Tag des zweiten Folgemonats) in das letzte Drittel der EU-Mitgliedstaaten. Die Sondervorauszahlung, die die Nutzer der Dauerfristverlängerung mit monatlicher

Tabelle 1
Meldefristen zur Abgabe der Umsatzsteuererklärung in Europa

	Umsatzsteuer-Voranmeldung	Zusammenfassende Meldung
	Tage nach Ablauf des Meldezeitraums	
Deutschland	10 ¹	25
Bulgarien	14	14
Luxemburg	15	25
Frankreich	19	10
Irland	19	23
Belgien	20	20
Estland	20	20
Kroatien	20	20
Lettland	20	20
Spanien	20	20
Ungarn	20	20
Dänemark	25	20
Litauen	25	25
Polen	25	25
Rumänien	25	15
Slowakei	25	25
Tschechische Republik	25	25
Griechenland	26	26
Schweden	26	25
Niederlande	30	30
Slowenien	30	20
Vereinigtes Königreich	37	21
Portugal	40	20
Zypern	40	15
Finnland	42	20
Malta	45	15
Österreich	45	30
Italien	60	25

1 Mit Dauerfristverlängerung: + 30 Tage.

Umsatzsteuer-Voranmeldung entrichten müssen, wird ausschließlich in Deutschland erhoben. [↪ Tabelle 1](#)

Aus unterschiedlichen Perspektiven haben sich bereits der Bundesrechnungshof, der Rechnungsprüfungsausschuss des Deutschen Bundestags und die Bundesländer mit der Problematik der auseinanderdriftenden Meldefristen beschäftigt. Auch in der Fachliteratur wurde diese Thematik eingehend behandelt mit dem Fazit, dass ein einheitlicher Termin für das gesamte Umsatzsteuergesetz geschaffen werden sollte. Die verschiedenen Termine zur Erklärungsabgabe stellten „schlicht eine überflüssige Bürokratie“ (Kemper, 2015, hier: Seite 378) dar. Das Bundesministerium der Finanzen begrüßt die Idee einer Vereinheitlichung der Termine, weist

jedoch auf die damit verbundenen haushalterischen Effekte hin. Durch den Wegfall der Sondervorauszahlung käme es im Zuge einer einmaligen Phasenverschiebung der Zahllasten im Jahr der Einführung zu einer Belastung der öffentlichen Haushalte in Milliardenhöhe, die in den Haushalten ausgeglichen werden müsse.

1.3 Projektziele

Das Bundeskanzleramt hat – in Abstimmung mit dem Bundesministerium der Finanzen und dem Bundesministerium für Wirtschaft und Energie – das Statistische Bundesamt beauftragt zu prüfen, wie sich eine Angleichung der Meldefristen zur Umsatzsteuer-Voranmeldung und zur Zusammenfassenden Meldung und der Meldezeiträume bei Wegfall der Dauerfristverlängerung auswirken. Neues Zieldatum der Abgabe sollte der letzte Kalendertag des auf den Meldezeitraum folgenden Monats sein. Im Fokus des Interesses standen die bürokratischen Belastungen der Unternehmen und der Verwaltung. Anhand einer Gegenüberstellung des Status quo mit dem Alternativszenario sollte das jeweilige Entlastungspotenzial aufgezeigt werden. Neben dem Erfüllungsaufwand für Wirtschaft und Verwaltung sollten auch Vor- und Nachteile erfragt und Verbesserungsvorschläge aufgenommen und strukturiert dargestellt werden.

Sowohl die Wirtschaft als auch die Verwaltung waren in dieser Untersuchung zu betrachten. Seitens der Wirtschaft stellte dies eine Herausforderung dar, weil – je nach Fallkonstellation – die Auswirkungen der Alternativszenarios in Gegensatz zueinander stehen können. So erhalten beispielsweise Unternehmen, die bislang die Dauerfristverlängerung nicht nutzen, eine Fristverlängerung von 20 Tagen. Dagegen hätten Unternehmen mit Dauerfristverlängerung künftig zehn Tage weniger Zeit, um ihre Umsatzsteuer-Voranmeldung an die Finanzverwaltung zu übermitteln. Je nachdem, ob darüber hinaus eine Meldeverpflichtung für die Zusammenfassende Meldung besteht, profitieren die Unternehmen von einer Harmonisierung der Fristen in unterschiedlichem Maße.

2

Methodik und Projektablauf

2.1 Erfüllungsaufwand

Das relevante Maß für Bürokratie und Grundlage der Messungen ist der Erfüllungsaufwand (Vorgrimler und andere, 2011). Er umfasst den Zeitaufwand und die Kosten, die den Bürgerinnen und Bürgern, der Wirtschaft und der Verwaltung durch die Befolgung einer gesetzlichen Vorgabe entstehen. Für die Berechnung des Personalaufwands werden die Zeitwerte für die Arbeitsschritte ermittelt, die zur Erfüllung einer gesetzlichen Vorgabe notwendig sind, und mit dem Lohnsatz der jeweils bearbeitenden Person multipliziert. Zum Sachaufwand zählen beispielsweise Kosten für die Inanspruchnahme externer Dienstleistungen (zum Beispiel Sachverständige), Kosten für Informations- und Kommunikationstechnik (zum Beispiel Spezialsoftware, Kartenlesegeräte) sowie Kosten für Sachmittel, wie Büromaterial oder Porto. Der Erfüllungsaufwand wird im Median ausgewiesen, da dieser robust ist gegen Ausreißer und sich somit am besten eignet, um ein normaleffizientes Verhalten abzubilden.

2.2 Datenerhebung

In der von August bis September 2018 andauernden Feldphase konnten Ansprechpartnerinnen oder Ansprechpartner aus 133 Unternehmen interviewt werden. Dazu wurden 468 zufällig ausgewählte Unternehmen telefonisch kontaktiert, um eine Person mit Fachexpertise für eine Gesprächsteilnahme zu gewinnen. Da ein Zugriff auf das statistische Unternehmensregister nach § 16 Bundesstatistikgesetz für dieses Projekt ausgeschlossen war, war es nicht möglich, die Stichprobe aus der Grundgesamtheit aller Normadressaten zu ziehen. Stattdessen wurden die Kontaktdaten aus dem Bestand einer kommerziellen Unternehmensdatenbank bezogen. Nach zuvor festgelegten Quoten (Branche, Unternehmensgrößenklasse, Bundesland), die auf einer Auswertung der Fachstatistik basierten, wurden 1 000 Kontaktdaten zufällig ausgewählt.

Die Stichprobe wurde nach einzelnen Fallkonstellationen (Unternehmen, die Dauerfristverlängerung nutzen, Unternehmen, die Zusammenfassende Meldungen abgeben) disproportional gewichtet. Dieses Vorgehen spiegelt sich in der Struktur der 133 befragten Unternehmen wider. Insgesamt wurden in Unternehmen aus exportorientierten Branchen (Handel, Verarbeitendes Gewerbe) aufgrund der Meldepflicht zur Zusammenfassenden Meldung deutlich mehr Interviews geführt als in Firmen aus dem Baugewerbe oder den persönlichen Dienstleistungen.

Bis auf Bremen – dort haben 0,7 % aller Unternehmen, die eine Umsatzsteuer-Voranmeldung abgeben, ihren Sitz – konnten aus allen Bundesländern Unternehmen für eine Teilnahme an den freiwilligen Interviews gewonnen werden. Erwartungsgemäß stammte rund die Hälfte der Befragten aus Bayern, Baden-Württemberg und Nordrhein-Westfalen. Die Teilnahmebereitschaft in Unternehmen aus den neuen Ländern und Berlin war unterdurchschnittlich. Obwohl rund 18 % der voranmeldepflichtigen Unternehmen ihren Standort in den neuen Ländern und Berlin hatten, kamen lediglich 13 % der Interviewteilnehmerinnen und Interviewteilnehmer von dort.

Für die Gespräche in den Unternehmen wurde ein eigens konzipierter Gesprächsleitfaden genutzt. In die Gestaltung des Fragebogens flossen Erkenntnisse aus Hintergrundgesprächen mit Verbänden ein. So konnte beispielsweise die Bundessteuerberaterkammer Auskunft über die typischen Arbeitsschritte bei der Bearbeitung einer Umsatzsteuer-Voranmeldung geben. Der Verband Deutscher Maschinen- und Anlagenbau hatte auf in der Praxis gehäuft auftretende Fehler beim Ausfüllen der Zusammenfassenden Meldung hingewiesen. Vor Beginn der Feldphase wurde der Gesprächsleitfaden getestet und abschließend überarbeitet. Dazu wurden in mehreren Unternehmen verschiedener Größenklassen und unterschiedlicher Branchen in mehreren Bundesländern persönliche Gespräche vor Ort geführt.

Der erste Teil des Gesprächsleitfadens enthielt allgemeine Fragen zum Unternehmen (Gründungsjahr, Umsatz, Beschäftigte). Anschließend wurde der Status quo abgefragt. Für die Umsatzsteuer-Voranmeldung und die Zusammenfassende Meldung waren dazu vornehmlich der jeweilige Abgabeterminus, die einzelnen Arbeitsschritte und ihr zeitlicher Aufwand, die Anzahl der not-

wendigen Korrekturen und das Qualifikationsniveau der Bearbeiterinnen und Bearbeiter von Bedeutung. Der Leitfaden umfasste auch Fragen, inwiefern ein Unternehmen die Möglichkeit einer Dauerfristverlängerung nutzt und welcher Aufwand mit der Anmeldung der dann fälligen jährlichen Sondervorauszahlung verbunden ist. Im zweiten Teil des Gesprächsleitfadens standen das neue Modell zur Harmonisierung der Umsatzsteuer-Meldefristen und die damit einhergehenden Vor- und Nachteile im Mittelpunkt. Zudem wurde ermittelt, ob sich durch das Modell einzelne Arbeitsschritte verändern, ein einmaliger Umstellungsaufwand anfällt und Synergien erwartet werden. Zum Ende der Interviews sollten die Unternehmen eine Einschätzung abgeben, ob sie dem neuen Modell zustimmend oder ablehnend gegenüberstehen.

Die Befragungen in den Unternehmen erfolgten telefonisch. In den Vorgesprächen wurde die jeweilige Fallkonstellation erfragt – beispielsweise monatliche Abgabe der Umsatzsteuer-Voranmeldung, Nutzung der Dauerfristverlängerung, Abgabe einer zusammenfassenden Meldung – sowie Branche und Umsatzgrößenklasse aufgenommen. Die Interviews selbst dauerten je nach Fallkonstellation zwischen 20 und 40 Minuten. Im Anschluss an das Interview wurden die Ergebnisse aufbereitet, dokumentiert und in einer Datenbank erfasst. Nach der Digitalisierung wurden die offenen Angaben der Befragten zusammengefasst und in vorgegebene Kategorien kodiert.

Um den Aufwand und die potenziellen Veränderungen für die Verwaltung zu erfassen, wurden mit dem Bundeszentralamt für Steuern und den Landesfinanzverwaltungen Expertengespräche geführt. Die Datenverarbeitung

erfolgt in diesen Behörden weitgehend automatisiert. Daher konnte sehr genau ausgewertet werden, wie häufig das System einen Bearbeitungshinweis generiert und manuell in den Bearbeitungsprozess eingegriffen werden muss. Auf Basis dieser Hinweise konnten Auswirkungen auf die Verfahrenskosten ermittelt werden.

3

Ergebnisse zum Status quo

3.1 Dauerfristverlängerung

Rund drei Viertel der Unternehmen, die zur Abgabe einer Umsatzsteuer-Voranmeldung verpflichtet sind, nutzen die nach § 46 UStDV bestehende Möglichkeit: Sie beantragten die Verlängerung der Abgabefrist um einen Monat. Eine Dauerfristverlängerung nutzten 74,1% der 1 293 000 Unternehmen, die ihre Umsatzsteuer-Voranmeldung vierteljährlich übermittelten. Der Anteil der Nutzer einer Dauerfristverlängerung unter den 1 973 000 Unternehmen, die monatlich ihre Umsatzsteuer-Voranmeldung abgeben müssen, ist ähnlich hoch: 72,6% dieser Unternehmen nehmen die Dauerfristverlängerung in Anspruch.¹ [↪ Grafik 1](#)

Die Unternehmen nutzen die Dauerfristverlängerung, um mehr Zeit bis zur Abgabe der Umsatzsteuer-Voranmeldung zu erhalten. Für den Wunsch nach mehr Zeit

¹ Siehe Sonderauswertung der Umsatzsteuerstatistik (Voranschlägen), Bezugsjahr 2016.

Grafik 1
Nutzung der Dauerfristverlängerung
in %



werden vornehmlich zwei Gründe genannt: Zum einen soll die Arbeitsbelastung entzerrt werden. Zum anderen sei die Buchhaltung bis zum Stichtag noch nicht abgeschlossen. Jeweils rund ein Viertel der Unternehmen nennt diese Gründe. Eine wesentlich geringere Rolle spielen noch zwei weitere Faktoren: Ein Teil der Unternehmen möchte durch die zeitlich nach hinten verschobene Abgabe der Umsatzsteuer-Voranmeldung spätere Korrekturen vermeiden, ein anderer Teil weist darauf hin, dass die Dauerfristverlängerung heute nicht viel mehr als ein „Überbleibsel aus der Vergangenheit“ sei.

Unternehmen, die von der Möglichkeit einer Dauerfristverlängerung keinen Gebrauch machen, begründen dies am häufigsten mit ihrer zur Frist passenden Arbeitsorganisation. Erst an zweiter Stelle wird als Grund für die Nichtnutzung der Dauerfristverlängerung ein erwarteter Vorsteuerüberhang genannt.

Die Unternehmen, die ihre Voranmeldungen vierteljährlich übermitteln, müssen nur einmalig eine Dauerfristverlängerung beantragen. Dazu reicht es, den einseitigen Antrag an einer Stelle mit einem zusätzlichen Kreuz zu versehen und an die Finanzverwaltung zu übermitteln. Aufgrund des geringen einmaligen Aufwands, der durch das zu prüfende Alternativszenario für die Unternehmen entfiel, wurde der Erfüllungsaufwand für diese Fallgruppe nicht ermittelt.

Monatsmelder müssen dagegen jährlich eine Sondervorauszahlung anmelden. Typischerweise müssen dafür fünf Arbeitsschritte absolviert werden: notwendige Daten beschaffen, Berechnungen durchführen, das Formular ausfüllen, die Daten und Eingaben überprüfen, archivieren. Der Zeitaufwand beträgt hierfür im Median 30 Minuten. Der durchschnittliche Stundenlohn beträgt 34,50 Euro. Daraus ergeben sich Kosten je Anmeldung einer Sondervorauszahlung in Höhe von 17,25 Euro. Wird diese Preiskomponente mit der Anzahl der Fälle je Jahr multipliziert, ergeben sich Gesamtkosten von 24,72 Millionen Euro für die Anmeldungen der Sondervorauszahlungen.

Die im Februar jedes Jahres insbesondere durch die Dauerfristverlängerung erheblich größere Anzahl an übermittelten Umsatzsteuer-Voranmeldungen beziehungsweise Sondervorauszahlungen führt nach Auskunft der Finanzverwaltung zu einer hohen Arbeitsspitze in den Finanzämtern. Zwar sei die Anmeldung der Sondervorauszahlung ein weitgehend automatisiertes Ver-

fahren, dennoch würde in rund 13 % der Fälle durch das System ein Prüfauftrag in Form einer Hinweisausgabe generiert. In rund der Hälfte dieser Fälle werden Anträge ausgesteuert, weil Unternehmen aktuell als steuerlich unzuverlässig eingestuft werden.

Obwohl demnach rund 87 % der Fälle automatisiert von der Verwaltung bearbeitet werden können, verursachen die ausgesteuerten Anträge einen hohen Zeitaufwand. Dieser ist so hoch, dass – bezogen auf alle Anmeldungen von Sondervorauszahlungen – die durchschnittliche Bearbeitungszeit je Fall durch die Finanzämter drei Minuten beträgt. In den Umsatzsteuer-Voranmeldestellen der Finanzämter erfolgt die Hinweisbearbeitung durch Beschäftigte des mittleren und gehobenen Dienstes. Daher wird ein durchschnittlicher Stundenlohn von 39,60 Euro für die Berechnung der Kosten herangezogen. Je Anmeldung einer Sondervorauszahlung ergeben sich danach Kosten von 1,98 Euro. Bei 1,433 Millionen Fällen betragen die Gesamtkosten für die Verwaltung im Sinne des Erfüllungsaufwands 2,84 Millionen Euro.

3.2 Zusammenfassende Meldung

Über die Anzahl der zur Zusammenfassenden Meldung meldepflichtigen Unternehmen wird keine Statistik geführt. Dem Bundeszentralamt für Steuern ist daher nicht bekannt, wie viele Unternehmen zur Übermittlung einer Zusammenfassenden Meldung verpflichtet sind. Deren Anzahl variiert in den Meldezeiträumen, da Unternehmen nicht zwingend fortlaufend am innergemeinschaftlichen Handel teilnehmen. Bei Nichtteilnahme ist in einem Meldezeitraum keine Null-Meldung abzugeben. Anhand der jährlich rund 1,5 Millionen einzelnen Zusammenfassenden Meldungen und der bekannten Exportquoten je Unternehmensgrößenklasse lässt sich die Anzahl der Unternehmen schätzen, die monatlich oder quartalsweise zur Abgabe einer Zusammenfassenden Meldung verpflichtet sind. Demnach sind rund 252 000 Unternehmen von dieser Meldepflicht betroffen.

Unternehmen, die zu einer monatlichen Abgabe der Umsatzsteuer-Voranmeldungen verpflichtet sind, müssen nicht zugleich zu einer monatlichen Abgabe der Zusammenfassenden Meldung verpflichtet sein. Für beide Meldungen gelten unterschiedliche Grenzbeträge. Ab einer jährlichen Zahllast von 7 500 Euro muss die Umsatzsteuer-Voranmeldung monatlich übermittelt

werden. Die Grenze zur Abgabe einer zusammenfassenden Meldung wird überschritten, wenn die innergemeinschaftlichen Warenlieferungen und Lieferungen im Sinne des § 25b Absatz 2 UStG im Rahmen von innergemeinschaftlichen Dreiecksgeschäften für das laufende Kalendervierteljahr oder für eines der vier vorangegangenen Kalendervierteljahre jeweils mehr als 50 000 Euro betragen. Somit treten die Entlastungspotenziale einer Harmonisierung der Meldefristen für die Umsatzsteuer-Voranmeldungen und die zusammenfassende Meldung je nach Meldezeitraum des Unternehmens unterschiedlich häufig auf. Für die weitere Betrachtung der Entlastungspotenziale wird der Anteil der Unternehmen, die monatlich eine zusammenfassende Meldung abgeben, aufgrund des Grenzbetrags von 50 000 Euro auf ein Drittel geschätzt.

Die Unternehmen wurden danach gefragt, welche Arbeitsschritte zusätzlich zur Bearbeitung der Umsatzsteuer-Voranmeldungen für die Abgabe der zusammenfassenden Meldung anfallen. Zunächst müssen die notwendigen Daten beschafft werden. Anschließend erfolgt eine Prüfung der korrekten Ziffernfolge der Umsatzsteuer-Identifikationsnummer. Je nach Umfang des Datenmaterials werden dann die Informationen in einer zusätzlichen Datei zusammengestellt oder in das Online-Formular eingetragen. Die Übermittlung erfolgt elektronisch an das Bundeszentralamt für Steuern. Für diese zusätzlichen Arbeitsschritte benötigen die Unternehmen im Median 30 Minuten Zeit. Bei einem Durchschnittslohnsatz von 34,50 Euro entstehen im Sinne des Erfüllungsaufwands Kosten von 17,25 Euro je zusammenfassender Meldung. Bei 1,495 Millionen jährlichen Meldepflichten belaufen sich die Gesamtkosten für die Wirtschaft auf 25,79 Millionen Euro.

Nur 58 % der Meldungen gehen beim Bundeszentralamt für Steuern bis zum gesetzlichen Abgabetermin ein. Die Konsequenz einer nicht fristgemäßen Abgabe ist ein automatisiertes Erzwingungsverfahren mit einer Erinnerung, einer Androhung und bis zu drei Zwangsgeldern. Daran schließt sich ein manuelles Bußgeldverfahren an. Ein automatisierter oder manueller Abgleich zwischen den Angaben aus der Umsatzsteuer-Voranmeldung und der zusammenfassenden Meldung findet durch das Bundeszentralamt für Steuern in Abstimmung mit den Ländern nicht statt. Ein solcher Abgleich wird wegen der unterschiedlichen Meldezeiträume als nicht zielführend betrachtet. Die Finanzämter können die Angaben

der Umsatzsteuer-Voranmeldung mit den Angaben zur Summe der Bemessungsgrundlagen in der zusammenfassenden Meldung überprüfen, soweit sie dies im Einzelfall für notwendig erachten. Hierzu können sie sich die Daten aus der zusammenfassenden Meldung mittels Zugriff auf die USLO-Datenbank („Umsatzsteuer Länder Online“) abrufen.

3.3 Korrekturen

Nachträgliche Korrekturen an bereits übermittelten Umsatzsteuer-Voranmeldungen oder zusammenfassenden Meldungen sind nicht selten. Der nachträgliche Korrekturbedarf tritt je nach Fallkonstellation unterschiedlich häufig auf. So geben 62 % der Unternehmen, die eine Dauerfristverlängerung nutzen und damit die zusammenfassende Meldung vor der Umsatzsteuer-Voranmeldung abgeben müssen, an, mindestens einmal eine bereits abgegebene zusammenfassende Meldung nachträglich korrigiert zu haben. Dagegen hatten nur 36 % der Unternehmen, die ihre Umsatzsteuer-Voranmeldung zum regulären Termin übermitteln und demnach ihre zusammenfassende Meldung bis zu 15 Tage nach der Umsatzsteuer-Voranmeldung fertigstellen können, nachträglichen Korrekturbedarf.

Am häufigsten verursachte nach Angaben der Unternehmen die Umsatzsteuer-Identifikationsnummer den Korrekturbedarf. So erweise sich beispielsweise die Umsatzsteuer-Identifikationsnummer eines ausländischen Geschäftspartners als nicht mehr gültig. Fehlerhafte oder unvollständige Rechnungen sind der zweithäufigste Grund für Korrekturen an bereits abgegebenen zusammenfassenden Meldungen.

Für die Korrektur einer zusammenfassenden Meldung wird genauso viel Zeit benötigt wie für deren Erstellung. Die Unternehmen benötigen 30 Minuten für diesen Vorgang. Jährlich fallen bei Unternehmen, die eine solche Berichtigung bereits einmal abgegeben haben und zugleich zu einer monatlichen Abgabe der zusammenfassenden Meldung verpflichtet sind, im Median 1,5 Korrekturmeldungen an. Quartalsmelder müssen durchschnittlich eine der vier Meldungen nachträglich abändern.

Ähnliches ist bei der Umsatzsteuer-Voranmeldung zu beobachten. Auch hier haben Unternehmen, die eine Dauerfristverlängerung nutzen und zugleich zur Abgabe

einer Zusammenfassenden Meldung verpflichtet sind, einen vergleichsweise hohen Korrekturbedarf. Rund 65 % der befragten Unternehmen aus dieser Fallgruppe mussten bereits mindestens eine Umsatzsteuer-Voranmeldung nachträglich korrigieren. Hingegen lag der Anteil bei Unternehmen, die zwar eine Dauerfristverlängerung nutzen, jedoch keine Zusammenfassende Meldung abgeben müssen, bei nur 35 %. Bei Unternehmen, die ihre Umsatzsteuer-Voranmeldung bereits zehn Tage nach Ablauf des Meldezeitraums übermitteln und zugleich keine Zusammenfassende Meldung abgeben müssen, stieg der Anteil auf 42 %. Rund 56 % der Befragten, die ihre Zusammenfassende Meldung nach der Umsatzsteuer-Voranmeldung abgaben, also keine Dauerfristverlängerung nutzten, mussten bereits mindestens einmal eine Umsatzsteuer-Voranmeldung korrigieren.

Der häufigste Grund für die nachträglichen Änderungen der Umsatzsteuer-Voranmeldung sind verspätet eingegangene oder gänzlich falsche Rechnungen. So gingen beispielsweise Rechnungen erst nach Abgabe der Umsatzsteuer-Voranmeldung ein, fehlten einzelne Rechnungsposten oder enthielten falsche Mehrwertsteuerangaben.

4

Entlastungspotenziale

4.1 Auswirkungen auf den Erfüllungsaufwand

Der Erfüllungsaufwand könnte sowohl für die Wirtschaft als auch für die Verwaltung durch eine Harmonisierung der Umsatzsteuer-Meldefristen spürbar gesenkt werden. Besonders ersichtlich wird dies durch den ersatzlosen Wegfall der Dauerfristverlängerung. Monatsmelder müssten künftig keine Sondervorauszahlung mehr anmelden. Wird diese Informationspflicht gestrichen, würde die bürokratische Belastung der Wirtschaft um 24,72 Millionen Euro gesenkt. Auch die bürokratische Belastung für Quartalsmelder, die einmalig einen Dauerfristverlängerungs-Antrag stellen müssen, könnte entfallen. Bislang sind dort für dieses Segment der Dauerfristverlängerungs-Nutzer rund 0,5 Millionen Euro verzeichnet.

Die Verwaltung könnte durch den Wegfall der Dauerfristverlängerung im Sinne des Erfüllungsaufwands um 2,84 Millionen Euro entlastet werden. Hinzu kommt die Vermeidung der unumgänglichen Arbeitsspitze im Monat Februar, die zu einer weiteren Entlastung der Verwaltung führen kann, da andere Anliegen nunmehr ohne zeitliche Verzögerung bearbeitet werden können. In der Praxis kommt es bislang beispielsweise zu einer verspäteten Bearbeitung von Erstattungswünschen im Monat Februar.

Durch eine Angleichung der Meldefristen und -wege sieht die Finanzverwaltung das Potenzial, die Datenqualität der Zusammenfassenden Meldung erheblich zu verbessern. Dadurch würde auch die Anzahl der Auskunftersuchen, die durch die EU-Mitgliedstaaten an Deutschland gestellt werden, zurückgehen und der damit verbundene personelle Aufwand gesenkt.

Eine höhere Datenqualität hätte auch Auswirkungen auf die Anzahl der Korrekturen bei Zusammenfassenden Meldungen und Umsatzsteuer-Voranmeldungen. Wird – wie bei den Nutzern der Dauerfristverlängerung vorgesehen – die Zusammenfassende Meldung künftig nicht vor der Umsatzsteuer-Voranmeldung erstellt und übermittelt, sondern zeitgleich mit dieser, sind nachträglich zu korrigierende Abweichungen zwischen beiden Meldungen durch eine Plausibilitätsprüfung in der Praxis nahezu ausgeschlossen. Der Anteil der Unternehmen, die mindestens einmal eine Korrektur bei einer Zusammenfassenden Meldung vornehmen müssen, beträgt derzeit 62 %. Es ist zu erwarten, dass er zumindest auf 36 % sinkt, also auf das Niveau, das heute für Unternehmen gilt, die ihre Zusammenfassende Meldung nach der Umsatzsteuer-Voranmeldung abgeben. Je nachdem, wie die Verknüpfung der beiden Meldungen technisch umgesetzt wird, könnten die bislang zusätzlich zur Bearbeitung der Umsatzsteuer-Voranmeldung notwendigen Arbeitsschritte für die Abgabe der Zusammenfassenden Meldung nahezu vollständig entfallen. Im Idealfall könnte die Wirtschaft dadurch um bis zu 25,79 Millionen Euro entlastet werden, das entspricht dem jetzigen Zusatzaufwand für die Zusammenfassende Meldung.

Ein Sechstel der Unternehmen, die bisher eine Zusammenfassende Meldung abgeben, erwartet durch den neuen Abgabetermin weniger Arbeitsschritte. Nach Ansicht der Befragten besteht die Möglichkeit, dass durch die Zusammenlegung und Angleichung der Melde-

zeiträume der Umsatzsteuer-Voranmeldung mit der Zusammenfassenden Meldung weniger Daten geprüft und eingegeben werden müssen. Die prognostizierte Zeitersparnis liegt hier im Median bei 30 Minuten.

4.2 Vor- und Nachteile aus Sicht der Unternehmen

Unternehmen, die zum regulären Termin die Umsatzsteuer-Voranmeldung abgeben und zusätzlich eine Zusammenfassende Meldung übermitteln müssen, hätten mehr Zeit für die Bearbeitung der Umsatzsteuer-Voranmeldung. Durch eine Verlängerung der Frist um 20 Tage würde Zeitdruck von den Bearbeiterinnen und Bearbeitern genommen. Rund ein Drittel dieser Unternehmen vermag keine Vorteile erkennen. Drei von vier Unternehmen aus dieser Fallkonstellation sehen keine Nachteile. Die restlichen Befragten vermuten vor allem in den neuen Arbeitsabläufen einen Nachteil für ihr Unternehmen. Auch um frühzeitig einen Überblick zu erhalten, sei es sinnvoll, wenn die Buchhaltung bis zum 10. des Folgemonats abgeschlossen sei. Ein weiterer Nachteil liege in der späteren Fälligkeit der Vorsteuerüberschuss-Rückerstattung, was auch einen einmaligen Cash-flow-Effekt mit sich bringe.

Unternehmen, die weder eine Dauerfristverlängerung nutzen noch eine Zusammenfassende Meldung abgeben, verbinden ganz überwiegend Vorteile mit dem angedachten Modell. Auch in dieser Gruppe wird positiv bewertet, dass künftig mehr Zeit für die Abgabe der Umsatzsteuer-Voranmeldung bestünde. In besonderen Situationen, beispielsweise bei Krankheit, Urlaub oder Dienstreisen der Mitarbeiterinnen und Mitarbeiter, entstünde durch die verlängerte Frist weniger Druck. Jedoch war etwas weniger als die Hälfte der Befragten der ausdrücklichen Auffassung, dass die neue Regelung keine Vorteile bringe. Vier von fünf Befragten sind dagegen der Meinung, die neue Regelung habe keine Nachteile.

Für Unternehmen, die bislang eine Dauerfristverlängerung nutzen und eine Zusammenfassende Meldung abgeben, stehen Vor- und Nachteile in einem anderen Verhältnis zueinander: Rund zwei Fünftel der Interviewten verbanden keine Vorteile mit dem neuen Modell. Ein Viertel sieht dem Wegfall der Sondervorauszahlung positiv entgegen. Zum einen müsse dann nicht mehr ein Elftel der verbleibenden Umsatzsteuer-Vorauszahlungen

des Vorjahres im Februar an die Finanzkasse überwiesen werden, zum anderen entfielen dann auch der mit der Anmeldung der Sondervorauszahlung verbundene Arbeitsaufwand. Negativ sticht aus Sicht dieser Interviewteilerinnen und -teilnehmer die verkürzte Meldefrist für die Umsatzsteuer-Voranmeldung hervor. Bemängelt wird, dass zehn Tage weniger zur Verfügung stünden, um die Meldung zu bearbeiten. Rund ein Drittel der Befragten aus dieser Fallkonstellation sah hingegen keine Nachteile.

Auch rund zwei Fünftel der Interviewteilerinnen und -teilnehmer, die eine Dauerfristverlängerung in Anspruch nehmen, jedoch keine Zusammenfassende Meldung abgeben müssen, sehen mit den neuen Meldefristen keine Vorteile verbunden. Von den restlichen Befragten dieser Fallkonstellation wird positiv hervorgehoben, dass künftig die Sondervorauszahlung entfallen soll. Fast die Hälfte dieser Unternehmen äußerte die Auffassung, durch die Harmonisierung der Fristen ergäben sich keine Nachteile. In jeweils unterschiedlichen Schattierungen bemängeln die anderen Gesprächspartnerinnen und Gesprächspartner dieser Konstellation wiederum die um zehn Tage verkürzte Bearbeitungszeit. Höherer Zeitdruck, geringere Flexibilität und mehr Korrekturen werden befürchtet.

5

Fazit und Ausblick

Seit 2010 unterscheiden sich in Deutschland nicht nur die Meldezeiträume für Umsatzsteuer-Voranmeldungen und Zusammenfassende Meldungen, sondern beide Meldungen müssen auch an unterschiedlichen Kalendertagen eingereicht werden. Dadurch ist die deutsche Rechtslage komplizierter geworden. Fraglich ist, ob das Ziel der Reform – eine bessere Bekämpfung des Steuerbetrugs bei innergemeinschaftlichen Umsätzen – erreicht wurde: Wegen der unterschiedlichen Abgabe- und Meldezeiträume wird kein automatisierter oder manueller Abgleich zwischen den Bemessungsgrundlagen der Umsatzsteuer-Voranmeldung und Zusammenfassenden Meldung durchgeführt. Darüber hinaus erfolgen vier von zehn Zusammenfassenden Meldungen zu spät.

Durch die Zusammenfassenden Meldungen entsteht der Wirtschaft zusätzlich zur Umsatzsteuer-Voranmeldung eine bürokratische Belastung von rund 25 Millionen Euro. Dieser Aufwand könnte durch eine Verknüpfung der Zusammenfassenden Meldung mit der Umsatzsteuer-Voranmeldung deutlich reduziert werden. Dadurch stiege auch die Datenqualität der Meldungen. Die Anzahl der verspäteten Übermittlungen, falschen Angaben und nachträglichen Korrekturen ginge zurück. Eine wirkungsvollere Bekämpfung des Steuerbetrugs wäre möglich.

Durch eine Harmonisierung der Meldefristen zur Umsatzsteuer und der Meldezeiträume könnte zugleich auch eine Besonderheit des deutschen Steuerrechts abgeschafft werden. Die Nutzung der Dauerfristverlängerung verursacht für die Wirtschaft einen jährlichen Erfüllungsaufwand von 24,72 Millionen Euro und für die Verwaltung von 2,84 Millionen Euro. Dem gegenüber steht eine jährliche Sondervorauszahlung der Unternehmen in Höhe von rund 17 Milliarden Euro (Stand Datenbasis: 2016). Diese wird im Regelfall mit der Umsatzsteuer-Voranmeldung des Monats Dezember wieder verrechnet. Die beschriebenen Maßnahmen waren im Rahmen des Projekts nicht zu bewerten oder bezüglich ihrer finanz- und haushaltspolitischen Folgen einzuordnen. Ob und inwieweit eine Harmonisierung von Meldefristen zur Umsatzsteuer die damit verbundenen Haushaltsminder-einnahmen rechtfertigen kann, bleibt einer gesonderten Prüfung vorbehalten. 

LITERATURVERZEICHNIS

Himmelsbach, Elke/Hornbach, Carsten/Michalik, Susanne/Kuehnhenrich, Daniel. [Methodische Grundlagen der Zufriedenheitsbefragungen zu behördlichen Dienstleistungen](#). In: WISTA Wirtschaft und Statistik. Ausgabe 4/2016, Seite 54 ff.

Kemper, Martin. *Erklärungspflichten des Umsatzsteuergesetzes – Kritische Betrachtung der zunehmenden Pflichten der Unternehmer im deutschen Umsatzsteuerrecht*. In: UR Umsatzsteuer-Rundschau. Jahrgang 64. Ausgabe 10/2015, Seite 373 ff.

Vorgrimler, Daniel/Bartsch, Gorja/Zipse, Christian. [Vom Standardkosten-Modell zur Messung des Erfüllungsaufwands](#). In: Wirtschaft und Statistik. Ausgabe 12/2011, Seite 1165 ff.

RECHTSGRUNDLAGEN

Umsatzsteuergesetz (UStG) in der Fassung der Bekanntmachung vom 21. Februar 2005 (BGBl. I Seite 386), das zuletzt durch Artikel 9 des Gesetzes vom 11. Dezember 2018 (BGBl. I Seite 2338) geändert worden ist.

Umsatzsteuer-Durchführungsverordnung (UStDV) in der Fassung der Bekanntmachung vom 21. Februar 2005 (BGBl. I Seite 434), die zuletzt durch Artikel 9 der Verordnung vom 12. Juli 2017 (BGBl. I Seite 2360) geändert worden ist.

Abgabenordnung (AO) in der Fassung der Bekanntmachung vom 1. Oktober 2002 (BGBl. I Seite 3866; 2003 I Seite 61), die zuletzt durch Artikel 15 des Gesetzes vom 18. Dezember 2018 (BGBl. I Seite 2369) geändert worden ist.

Gesetz über die Statistik für Bundeszwecke (Bundesstatistikgesetz – BStatG) in der Fassung der Bekanntmachung vom 20. Oktober 2016 (BGBl. I Seite 2394), das zuletzt durch Artikel 10 Absatz 5 des Gesetzes vom 30. Oktober 2017 (BGBl. I Seite 3618) geändert worden ist.

Herausgeber

Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung

Dr. Daniel Vorgrimler
Redaktionsleitung: Juliane Gude
Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

zweimonatlich, erschienen im August 2019
Das Archiv älterer Ausgaben finden Sie unter www.destatis.de

Print

Einzelpreis: EUR 19,- (zzgl. Versand)
Jahresbezugspreis: EUR 114,- (zzgl. Versand)
Bestellnummer: 1010200-19004-1
ISSN 0043-6143
ISBN 978-3-8246-1083-9

Download (PDF)

Artikelnummer: 1010200-19004-4, ISSN 1619-2907

Vertriebspartner

IBRo Versandservice GmbH
Bereich Statistisches Bundesamt
Kastanienweg 1
D-18184 Roggentin
Telefon: +49 (0) 382 04 / 6 65 43
Telefax: +49 (0) 382 04 / 6 69 19
destatis@ibro.de

Papier: Design Offset, FSC-zertifiziert

© Statistisches Bundesamt (Destatis), 2019

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.